# DOCUMENT RESUME

ED 192 331 CS 205 845

| | |
|---|---|
| AUTHOR | Felker, Daniel B., Ed. |
| TITLE | Document Design: A Review of the Relevant Research. |
| INSTITUTION | American Institutes for Research, Washington, D.C.; Carnegie-Mellon Univ., Pittsburgh, Pa.; Siegel & Gale, Inc., New York, N.Y. |
| SPONS AGENCY | National Inst. of Education (DHEW), Washington, D.C. |
| REPORT NO | AIR-75002-4/80-TR |
| PUB DATE | Apr 80 |
| CONTRACT | 400-78-0043 |
| NOTE | 171p. |
| EDRS PRICE | MF01/PC07 Plus Postage. |
| DESCRIPTORS | *Design Requirements; *Language Processing; Layout (Publications); Literature Reviews; *Publications; *Readability; *Reading Research; State of the Art Reviews; *Written Language |
| IDENTIFIERS | *Document Design Project |

ABSTRACT

Research from several disciplines has been examined to create this literature review of information on document design, the overall movement toward producing public documents that the intended users can understand. Six chapters review appropriate research from the areas of psycholinguistics, cognitive psychology, instructional research, readability, human factors, and typography/graphics. A seventh chapter, written in the form of a case study, illustrates the development of an actual document design experiment, showing how research from the different disciplines can be used collectively to treat a typical document design issue. An appendix contains lists of references used in the seven chapters. (RL)

***********************************************************************
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
***********************************************************************

ED192331

AIR-75002-4/80-TR

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

DOCUMENT
DESIGN
PROJECT

# Document Design: A Review of the Relevant Research

Daniel B. Felker, Editor
American Institutes for Research

With chapters by:

Marshall Atlas
Carnegie-Mellon University

Veda R. Charrow
American Institutes for Research

Daniel B. Felker
American Institutes for Research

V. Melissa Holland
American Institutes for Research

Cheryl Olkes
Siegel & Gale, Inc.

Janice C. Redish
American Institutes for Research

Andrew M. Rose
American Institutes for Research

A product of the Document Design Project funded by
the National Institute of Education, DHEW, Washington, DC
Contract No. NIE-400-78-0043

April 1980

AMERICAN INSTITUTES FOR RESEARCH/1055 Thomas Jefferson Street, NW, Washington, DC 20007 • 202/342-5000
with Carnegie-Mellon University and Siegel & Gale, Inc.

CS205845

OCT 2 0 1980

This report is part of the work being done under contract No. 400-78-0043 of the National Institute of Education. The project officer is Dr. Ramsay Selden of the Teaching and Learning/Reading and Language Group. This report does not, however, necessarily reflect the views of the funding agency.

# ABOUT THE DOCUMENT DESIGN PROJECT

In September 1978, the American Institutes for Research (AIR) began the Document Design Project to foster clear and simple writing and design of public documents. The purpose of the Document Design Project (DDP) is to help make forms, regulations, brochures, and other written materials easier for people to read, to understand, and to use.

Carnegie-Mellon University and Siegel & Gale, Inc., a private firm that specializes in language simplification and language training, work with AIR on this project. Funding for the project comes from the National Institute of Education.

The Project's goal is to increase the knowledge and skills of people who produce public documents. To accomplish this goal, the Document Design Project is working on three tasks:

Task 1: To conduct theoretical and applied research studies on language comprehension, on the ways in which skilled and unskilled writers work, on problems associated with different document features.

Task 2: To work directly with government and private agencies as they produce materials for public use.

Task 3: To develop courses on writing and design for graduate students and undergraduates.

If you have questions or comments on this report or on other work of the Document Design Project, contact Dr. Janice C. Redish, Director, The Document Design Project, AIR, (202) 342-5071.

# NOTE ON AUTHORSHIP

This report reflects the combined effort of the Document Design Project staff at the American Institutes for Research, Siegel & Gale, and Carnegie-Mellon University. One person was responsible for each chapter, but wrote it with assistance from other staff. The purpose of each chapter was to describe the relationship between various research disciplines and document design and to demonstrate the interrelationship among the disciplines. This required that the staff negotiate, debate, and renegotiate the content in order to draw the line between what parts of the discipline relate to document design and under which discipline certain research should be placed. The authors responsible for each chapter are:

- Psycholinguistics (Chapter I) by Dr. Veda Charrow with assistance from V. Melissa Holland.
- Cognitive Psychology (Chapter II) by Dr. Marshall Atlas, with assistance from Dr. John Hayes, Dr. Lee Gregg, and Dr. Andrew Rose.
- Instructional Research (Chapter III) by Dr. Daniel Felker.
- Readability (Chapter IV) by Dr. Janice C. Redish.
- Human Factors (Chapter V) by Dr. Andrew Rose, with assistance from Dr. Lee Gregg.
- Typography (Chapter VI) by Dr. Cheryl Olkes.
- A Case Study: Complex Instructions (Chapter VII) from a study by V. Melissa Holland and Dr. Andrew Rose.

In addition, much valuable assistance was provided by Laura Malakoff, who checked references and helped coordinate the various efforts. Edith Griffin provided excellent and consistent typing support.

D.B.F.

# Table of Contents

page

INTRODUCTION . . . . . . . . . . . . . . . . . . . . . . . . . 1

CHAPTER I: Psycholinguistics . . . . . . . . . . . . . . . . . 7

Words . . . . . . . . . . . . . . . . . . . . . . . . 8
Sentences . . . . . . . . . . . . . . . . . . . . . . 12
Context and Discourse Structure . . . . . . . . . . . . 21

CHAPTER II: Cognitive Psychology . . . . . . . . . . . . . . 27

Variety . . . . . . . . . . . . . . . . . . . . . . . 30
Flexibility . . . . . . . . . . . . . . . . . . . . . 31
Hierarchical Structure . . . . . . . . . . . . . . . . 32
Use of Non-Linguistic Information . . . . . . . . . . . 34
Use of Procedural Information . . . . . . . . . . . . 36

CHAPTER III: Instructional Research . . . . . . . . . . . . 41

Programmed Instruction . . . . . . . . . . . . . . . . 43
Prose Learning . . . . . . . . . . . . . . . . . . . . 46
Prose Text Manipulations . . . . . . . . . . . . . . . 56
Aptitude-Treatment Interaction . . . . . . . . . . . . 58
Instructional Technology . . . . . . . . . . . . . . . 62

CHAPTER IV: Readability . . . . . . . . . . . . . . . . . . . 69

Introduction . . . . . . . . . . . . . . . . . . . . . 69
History of Readability Research . . . . . . . . . . . . 71
Issues in the Study of Readability . . . . . . . . . . 84

CHAPTER V: Human Factors . . . . . . . . . . . . . . . . . . 95

*HEGED* . . . . . . . . . . . . . . . . . . . . . . . 96
Other Research . . . . . . . . . . . . . . . . . . . . 99
In Conclusion . . . . . . . . . . . . . . . . . . . . 101

CHAPTER VI: Typography/Graphics . . . . . . . . . . . . . . 103

Typography . . . . . . . . . . . . . . . . . . . . . . 104
Graphic Design . . . . . . . . . . . . . . . . . . . . 106

CHAPTER VII: A Case Study: Complex Instructions . . . . . . 111

Understanding Complex Conditional Instructions . . . . 112

APPENDIX: References

Psycholinguistics . . . . . . . . . . . . . . . . . . . 125
Cognitive Psychology . . . . . . . . . . . . . . . . . 139
Instructional Research . . . . . . . . . . . . . . . . 143
Readability . . . . . . . . . . . . . . . . . . . . . 153
Human Factors . . . . . . . . . . . . . . . . . . . . 159
Typography/Graphics . . . . . . . . . . . . . . . . . 163
A Case Study: Complex Instructions . . . . . . . . . . 167

# Introduction

Every day, millions of people read public documents prepared by government agencies, commercial and financial organizations. These public documents include regulations for every conceivable kind of product and process, labels for over-the-counter and prescription drugs, rental agreements and insurance policies, and income tax instructions and forms. Public documents are set apart from other kinds of printed and graphic materials in two significant ways: 1) they convey information that directly affects people's social and physical well-being, and 2) a large proportion of them are barely comprehensible.

The fact that public documents are dull and hard to understand is not new. Legalese, bureaucratese, and gobbledygook have been with us a long time. What is new is the recognition in official circles that the dismal state of public documents is neither necessary nor desirable. Several state legislatures recently have passed laws that require contracts to be written in language that the average citizen can understand. Consumer groups are pushing for rules that ensure that information affecting consumers is clear. A Presidential Executive Order requires federal agencies to review their regulations for clarity. Some banks and insurance companies are rewriting loan agreements, policies, and other documents in "plain English."

We use the term "document design" to describe the overall movement aimed at producing effective public documents--documents that are comprehensible to their intended users. Document design is just now emerging as a recognizable field of practice and a legitimate area of research. The current boundaries of the field are hazy and it is uncertain what the field will look like in the future. It is certain that it has behind it the force of the public increasingly demanding clearer documents, a demand likely to be around for a long time. This is sufficient reason to begin to describe the field and its research domain in more detail.

The words "document design" are a deliberate choice because they convey the complexity of the field. The terms "clear writing" and "plain English" are not sufficient because useful, understandable documents entail more than easy words and simple sentences. The organization and format of a document may be just as important as its language. The degree to which the document is matched to the capabilities of its users and the setting of its use may affect comprehension as much as clearly written sentences. The broader term "document design" encompasses these added complexities.

But what is a clear document? How does one know when a document is comprehensible for a given set of users? What does a document designer do to make a document understandable for audience "X"? For audience "Y"? These questions and many more besides cannot be answered at this time. There is a conventional wisdom based on accumulated practical experience, but no generalizable knowledge derived from systematic research.

If document design is to mature as a field, and if document designers ever are going to be able to improve public documents on a significant scale, research will have to help guide the way. At present there is no existing research literature that pertains specifically to document design. There is, however, a considerable body of research from other disciplines that relates to document design. It is the research from these other disciplines that can serve as starting points for developing a document design research literature, that will help define the appropriate research boundaries, and that will uncover research priorities. The need is to identify the existing bits and pieces of research and, ultimately, to synthesize the research into some kind of coherent document design structure.

## PURPOSE AND APPROACH

Our overall goal in this review is to begin to develop a document design research literature. Our approach has been to examine research from different disciplines that relate to document design. In addition, we have several intermediate objectives that ultimately will help achieve the longer range goal. These are:

- to pull together existing research from a variety of disciplines and place it in a document design context;
- to illustrate the multi-disciplinary basis and potential of document design research;
- to demonstrate some scientific theories and concepts and some research methods and techniques that are applicable to document design issues;
- to stimulate researchers from different disciplines to think about issues in document design;
- to point out shortcomings and pitfalls experienced in other research disciplines as cautions for document design researchers;
- to suggest areas in document design where practitioners and researchers can work together and profit from each other.

We view our literature review as a start. It is not exhaustive. Out of necessity we have been selective--in the disciplines we chose to review, in the research areas within each discipline we included, and in the specific studies within each research area we cited.

There are many different rationales for selecting the particular research to include in literature reviews. Our rationale was straightforward and was based on the key components involved in any public document situation. The first of these components is the document itself, which necessarily has certain language, format, and organizational features. The second component is the intended user of the document, the person who is expected to read or use the document for some purpose. The third component is the document designer who actually writes the words, does the graphics, or makes decisions about the final product.

Given this approach, the proper research to include in a document design literature review obviously is that which says something about these three components. It follows that we had to be concerned with research that examined relationships between document features and the comprehension of documents. For example, what is it about language that makes a prose passage clear or ambiguous? When can illustrations be more informative than words? Do variations in color, or column widths, or type of print make any difference in comprehension? We also had to think about research that deals with users of documents. How do humans process information contained in documents? What factors facilitate the encoding, storage, and retrieval of document content? What individual differences in users are significant for understanding public documents? How can we measure whether a user really understands a document? Similarly, we had to consider the people who actually produce and design public documents. What are the essential skills for document designers? How can substantive experts best be trained to communicate clearly in documents? What processes and techniques should be used for developing and testing public documents?

We settled on six research disciplines which, in our opinion, encompass a wide array of issues bearing on document features, document users, or document designers. The disciplines and our reasons for selecting them are:

- psycholinguistics, because the emphasis on the comprehension, memory, and use of written and spoken language provides insights into what makes public documents easy or hard to understand and use;
- cognitive psychology, because cognitive theory provides the crucial underpinning for integrating the findings from many academic disciplines;
- instructional research, because the factors which influence how one learns from instructional materials and texts have similarities with factors which influence how one comprehends public documents;

- readability, because readability formulas are currently the standard that is used to measure how well a document matches the reading level of its audience;
- human factors, because its focus on direct, practical human-environment interactions includes studies on many document features;
- typography, because it deals with the basic features of documents that can be manipulated in numerous ways.

In addition, we have included a chapter in the form of a case study that illustrates the development of an actual document design experiment. The case study is instructive in that it shows how research from different disciplines can be pulled together in examining a typical document design issue. Even though we present essentially descriptive research reviews here, the case study makes the fundamental point that document design is a multi-disciplinary, integrated undertaking.

Within each discipline, we have concentrated on studies that have direct relevance to designing public documents. We primarily considered research done with adult readers or learners (high school and above). We focused on research that involved meaningful prose of sentence length or greater. We stressed research that required memory, recall, and comprehension tasks because public documents require similar information processing activities.

# I. Psycholinguistics

Much of psycholinguistic research deals with people's ability to understand and use their native language, in spoken or written form. As such, the research seeks to provide information on what is involved in language comprehension, on how people process and produce written and spoken language, and on what cognitive processes influence how they encode and decode words, sentences and discourse.

Psycholinguistic research can be valuable to document design because it can provide empirical support for the principles of (written) language use advocated by document design experts. But as the reader will note in the following literature review, there are gaps in the literature--areas which have been overlooked or only sketchily investigated by psycholinguists. Many guidelines for clear writing have little or no empirical support--although they have not been disconfirmed either.

In this chapter, we review the literature in the area of psycholinguistics. Many of the investigations of language processing are based on models of cognitive processes, on theories and information about human memory and how it works, and on the effects of individual differences. We will refer to such models, theories, and results in our discussion of studies of the various aspects of language and thought.

We could have organized this literature review from the perspective of the various psycholinguistic theories and models; however, since the theories and models are often disputed, or their validity questioned on the basis of alleged methodological flaws or inconsistent results, we decided that an organization based on traditional linguistic categories would be more straightforward and less subject to disputes. We have therefore organized this review in the following way:

Words. This section will deal with psycholinguistic research into the comprehension of individual words and word lists. We will cover such aspects as word frequency, markedness, and abstract versus concrete words and their relation to comprehension and comprehensibility. We also will deal with the beginnings of syntax--units larger than one word but smaller than a clause--noun strings, phrases, and sentence fragments.

Sentences. In this section, we will cover research involving complexity at the sentence level--divided roughly into two subsections: a) grammatical complexity, and b) semantics and pragmatics. The bulk of the research in recent years has focused on various aspects of semantic processing, presupposition, inference, and context.

Discourse. This section will cover research into linguistic processing of units larger than a sentence. We will deal with integration of information from lists of sentences, and with the comprehension and comprehensibility of entire passages. Most of the focus of this research has been semantic and pragmatic, although some has related to grammatical considerations, as well.

Throughout our discussion, we will note those studies that have direct relevance to problems of comprehension in the real world, and that provide insight into ways of producing clearer documents. For somewhat different perspectives on the same field, and additional references, see Danks and Glucksberg's review of the literature in experimental psycholinguistics, and Wright (1978) and Reder (1978).

## WORDS

Research into the processing of words has studied such dimensions as frequency, length, pronounceability, imagery, concreteness, meaningfulness (the number of associations with all other words), and strength of associations between words within a learning list (Noble, 1952; Deese, 1962; Palermo & Jenkins, 1964; Paivio, 1969; Cox, 1978; Scarborough, Gerard & Cortese, 1979). Results have shown in general that the shorter the word, and the higher it rates on any of the other dimensions, the more easily it is learned. These results are generally valid whether the test is serial, paired associate, or free recall learning (Hall, 1954; Cieutat, Stockwell & Noble, 1958; Jenkins, Mink & Russell, 1958; Underwood & Schultz, 1960; Rothkopf & Coke, 1961; Gibson, et al., 1964; Deese, 1965). These dimensions have the same effect on speed of recognizing a word in procedures that do not require learning (Howes & Solomon, 1951; Howes, 1957; Gibson, et al., 1964).

Studies have been conducted that show a strong positive relationship between the frequency of a word and its meaningfulness (Loewenthal, 1969; Klare, 1968). There have also been studies of specific classes of high and low frequency words which indicate that more frequent items are in fact better understood (Noble, 1952; O'Neill, 1972--nouns; Smith & Dixon, 1971--past tense verbs; Hogenraad, 1969--adjectives; Carroll, 1971--a mixture).

Other studies support these results (Solomon & Postman, 1952; Brown & Rubenstein, 1961; Broadbent, 1967; Champagnol, 1971; Klintman, 1968; Morton, 1968; Paivio & O'Neill, 1970). But Geis and Winograd (1975) have shown that different contexts will bias subjects' judgments of the relative frequency of homographs (words which are spelled the same), and Charrow and Charrow (1978, 1979 a & b) found that for some specific low frequency legal terms in the context of jury instructions, subjects did not have the expected difficulties. These studies give some indication of the importance of context for understanding low frequency words.

Studies of abstract and concrete words have been particularly useful (Paivio, 1969; Cox, 1978) in that they have served as a basis for investigations of the processing of concrete and abstract sentences (see following section). Katz and Denny (1977) have demonstrated that even when words are equivalent in frequency and meaningfulness, those embodying more concrete concepts will be more easily learned and recalled. All of this research argues for the use of concrete versus abstract words in writing documents. Creating concrete examples of abstract concepts would be useful in those cases where abstract ideas cannot be avoided.

Cofer (1961), Cofer & Musgrave (1963), Mandler (1967), and Dixon & Horton (1968) among others have reviewed the effects of word variables, the techniques used to measure them, and the findings and implications of verbal learning research in general.

A relatively small amount of psycholinguistic research has focused on units larger than a word, but smaller than a clause. Such studies have involved, among other things, noun strings, the effect of knowing what a verb means on memory for adjectives, and the effect of single word context on the meanings of homonyms. Thus, a study by Gleitman and Gleitman (1970) supports the notion that strings of nouns, in which there are few clues to syntactic structure, are often difficult to understand. Ehri and Muzio (1974), in a study of memory for subject- or object-modifying adjectives, found that adjectives that were semantically unrelated to verbs were recalled better than adjectives that were semantically related to verbs. The very unpredictability of the unrelated adjectives probably made them more memorable.

In verbal learning studies, words are usually presented to the subject in isolation, with no natural language context as there would be in an actual document. But some studies have been performed which place lexical items in a sentence context. Although some of these do in fact deal with the syntax and semantics of the *sentence*, others are really simply studies of *word* effects.

Begg and Clark (1975) used sentence fragments to emphasize the two meanings of homonyms. They found that recall of each meaning of the word out of context could be predicted as a function of the relative frequency for that meaning and a rating of the image-arousing capacity of that meaning in context. Recall of words presented in context was a function of contextual imagery. Similarly, Halff, Ortony and Anderson (1976) found that subjects' interpretation of the color meant by the adjective "red" varied on a continuum, depending on sentence context (e.g., fire engine red vs. sunburn red).

Warren, Warren, Green and Bresnick (1978), using single words as a context to disambiguate homophones, found, as had Cairns and Kamerman (1975), that all meanings of ambiguous words appear to be activated during encoding. However, in contrast to Cairns and Kamerman, Warren et al. found that even after one meaning has been selected, the other meanings remain active in memory for a long period of time.

On the other hand, there is a type of study of sentence effects which may actually be measuring only word effects. Thus, in studies of abstract and concrete sentences, the abstractness or concreteness of a sentence is gauged entirely by the abstractness or concreteness of the words it contains. Results from such studies (Begg & Paivio, 1969; Johnson, Bransford, Nyberg & Cleary, 1972; Franks & Bransford, 1972; Pezdek & Royer, 1974; Kuiper & Paivio, 1977; Green, 1977), tend to confirm the findings regarding abstract and concrete words in isolation: concrete sentences are more easily understood than abstract ones. However, a few such studies also show that there is more to understanding sentences than simply understanding their component words (Hogaboam & Perfetti, 1975; Fischler & Bloom, 1979).

Although some studies have shown that sentence context effects override memory for specific wording (Begg & Paivio, 1969), others show that lexical (word) information persists in memory representations of meaning (Hayes-Roth & Hayes-Roth, 1977). This particular line of research has a parallel in investigations of sentence processing, notably, in studies of whether and how syntactic information about sentences is stored and retrieved. We will deal with this in the next section.

# SENTENCES

Until a few years ago, most psycholinguistic research involved either words in isolation, or sentence processing. Consequently, there is an enormous amount of data involving numerous aspects of sentence comprehension and recall, and many models of sentence processing. Our discussion of psycholinguistic studies involving sentences is divided roughly into studies of grammar, and studies of semantics (including context and pragmatics).

## Grammar

Psycholinguistics got a new impetus--even rebirth--with the development of transformational grammar (TG) as an explanation of language structure (Chomsky, 1957; 1965). From 1960 to early in the 70s, the field of psycholinguistics was dominated by studies based on Chomsky's theories and by work combining transformational theory and developments in cognitive psychology. Fillenbaum (1971), in the *Annual Review of Psychology*, surveys the first decade of research based on the new theory. Bever (1968), Garrett and Fodor (1968), and Wales and Marshall (1966) give more detailed reviews of some of the same material. Early studies compared basic sentence types, effects of single or increasing numbers of transformations, and effects of complicating underlying structure. Experimental psychologists attempted to adopt TG as a model of actual sentence processing, and early studies tried to demonstrate the psychological reality of various TG constructs (e.g., Miller, 1962, 1964).

As a model of human sentence processing, transformational grammar offered a precise metric for predicting the relative difficulty of sentences. According to TG, the structure of the simple active affirmative declarative (kernel) sentence* is the basis from which all other sentence types are derived through transformations. Thus, according to TG, the kernel is psychologically simplest and each transformation adds to the complexity of the sentence.

---

*E.g., Mary kissed John.

The metric worked in early studies of relative processing difficulty. These studies compared four sentence types in various combinations--the kernel (K); the transforms: question (Q); negative (N); passive (P); and the combinations QN, NQ, PN, PQ, PQN.** Processing difficulty was measured by testing memory or comprehension. The order of difficulty can be represented:

$$\begin{array}{c} PQ > P \\ PQN > NQ > Q > K \\ PN > N \end{array}$$

Numerous studies appeared to support this TG model: Miller (1962), Mehler (1963), Mehler and Miller (1964), Miller and McKean (1964), Savin and Perchonock (1965), Clifton, Kurcz and Jenkins (1965), Clifton and Odom (1966), Marshall (unpublished), and Gough (1965, 1966).

In other studies of transformational complexity, researchers found the following results, which can also be of use to document designers:

- Subjects take longer to complete sentence verification tasks when the sentence is negative than when it is affirmative--e.g., "7 is not an even number" (Wason, 1961; Eifermann, 1961).
- Negative qualifiers (e.g., "except") impede response rate (Jones, 1966).
- The simple sentence is easier to understand than when it is transformed into nominalized form--e.g., "Mary searched for John" rather than "Mary's searching for John" (Coleman, 1964).
- Sentences with relative clauses are easier to understand when the relative pronoun is retained than when it is deleted--e.g., "He said things were going well" vs. "He said *that* things were going well." (Fodor & Garrett, 1967; Hakes & Cairns, 1970).

---

**e.g., Wasn't John kissed by Mary? (PQN)

- Complex sentences with the subordinate clause first are more difficult to recall than the basic main-clause-plus-subordinate-clause structure (Clark & Clark, 1968).
- Embedding a clause inside the sentence (self- or center-embedding) rather than stringing it on the end (right-branching) makes comprehension and recall more difficult (Miller, 1962; Miller & Isard, 1964; Stolz, 1967; Schwartz, Sparkman, & Deese, 1970).

Despite its apparent dominance of the field, TG was not the only model being used in psycholinguistic research, even in the 1960s. Another approach was based on phrase structure grammar. Johnson (1965, 1968) measured the functional units for understanding a sentence in terms of word-to-word transitional error probabilities (TEPs) and found these units were identical with phrase structure constituents. Schlesinger (1968) supported this result and suggested, as had North and Jenkins (1951) and Coleman and Kim (1961), that reading can be speeded up by dividing the line of print into phrase structure units.

Yngve (1960) defined sentence complexity in terms of how deeply phrases are embedded and where they occur in the sentence. For example, nesting (embedding) at the beginning is more complex than at the end. Yngve called this measure of complexity "sentence depth." He hypothesized that greater depth makes a sentence harder to understand and produce because it places a greater load on short-term memory (STM). Tests of the depth hypothesis have produced mixed results. It may be that semantic factors interact to affect sentence difficulty. Martin and Roberts (1967) and Martin, Roberts, and Collins (1968) found that depth overrides transformational complexity, but studies by Rohrman (1968), Wright (1960), and Perfetti (1966) failed to find significant effects due to depth.

The early enthusiasm of psychologists for the transformational approach waned as later studies failed to replicate those findings. For example, the Savin and Perchonock (1965) result of STM difficulty following transformational complexity did not hold up in replications by Matthews (1968), Glucksberg and Danks (1969), and Epstein (1969). Turner and Rommetveit (1968) found counter-transformational evidence using Mehler's (1963) procedure. In addition, more recent studies of the passive voice have shown that where passive focus is necessary, the passive may actually be easier to process than its active counterpart (Huttenlocher & Strauss, 1968; Huttenlocher & Weiner, 1971; Hornby, 1974; Hupet & Le Bouedec, 1975; Charrow & Charrow, 1979). In such cases, Wannemacher (1974) found that the passive is no more difficult to process than the active.

There have been other objections to psycholinguistic studies based on TG. Fillenbaum (1970) questions the validity of those studies using memory techniques to measure syntactic difficulty. Clark and Clark (1968), in their study of memory for complex sentences, assert that memory for sentences has a semantic, rather than syntactic, basis. There apparently are confounding semantic, pragmatic, and performance factors that make the real-world situation much more complex than the earlier studies suggested.

All this appears to have put a damper on studies of the relation of grammatical complexity to comprehension (whether or not the complexity is measured in TG terms), with the result that effects of grammatical complexity on language processing have been largely ignored in recent years. The few studies that have been performed have not been based on a TG model; other theories have been used to explain their results. But whether or not one accepts the theoretical bases of TG, the findings from many of the above studies are of interest to document designers. Most can be replicated, and findings such as those concerning the facilitating effect of relative pronouns (Foss & Garret, 1967; Hakes & Cairns, 1970), non-nominalized sentences, and right-branching versus center-embedding are still valid and can be applied by document designers.

A few more recent investigations of grammatical complexity continue to show that in spite of semantic, contextual, pragmatic and performance factors, grammatical complexity does play a significant role in the processing of language. Danks and Sorce (1973) have shown that full passives are recalled better than truncated passives. The Charrow and Charrow (1979) investigation of jury instructions revealed that passives in subordinate clauses (often accompanied by whiz deletion--deletion of relative-pronoun-plus-copula [BE verb]) are far more difficult to comprehend and remember than either full or truncated passives in main clauses. A study by Hakes, Evans and Brannon (1976) compared the comprehension of sentences with one center-embedding versus sentences with one right-branch, to test the hypothesis that center embeddings are difficult to process because such clauses interrupt the clauses in which they are embedded. Their results consistently failed to support the interruption hypothesis; however, other studies (e.g., Larkin & Burns, 1977) have shown that, although a single center-embedding causes relatively few problems, comprehension does decline as a function of the number of center embeddings. It is interesting to note that Larkin and Burns speak of this effect in terms of short-term memory overload, rather than grammatical complexity, but "grammatical complexity" may in fact be a specialized aspect of more general problems in cognitive processing (Bever, 1970). Such studies point out the need for more interaction between researchers in cognitive psychology and theoretical linguistics.

The notion that short-term memory plays a part in the poorer comprehension of grammatically complex sentences is supported in a study by Richek (1976). The experimenter found that children's reading comprehension of complex sentences was worse if the subordinate clause separated the subject from the verb phrase than if it followed the verb phrase. These effects were attributed to a short-term memory effect.

In their investigation of the comprehensibility of standard jury instructions, Charrow and Charrow (1978, 1979 a & b) isolated a number of purely grammatical features of the jury instructions that caused comprehension problems for the subjects. These included nominalizations, subordinate-clause passives (with whiz-deletion), deletion of relative pronouns, multiple negatives, and misplaced phrases. In addition, effects were found for sentences with unusual subordinate clause embeddings, which were apparently not simply a function of the number of subordinate clauses. While there were semantic, contextual, pragmatic and discourse features in the jury instructions that also created comprehension problems, it was clear that various types of grammatical complexity significantly affected subjects' comprehension.

As noted previously, there has been a whole area of psycholinguistic research devoted to memory for syntactic form as well as semantic aspects of sentences--sometimes explained in TG terms of memory for surface structure and deep structure. These studies are important for theories of human understanding in that they provide information about how people "get the gist" of what is said or written (see also Fillenbaum, 1966). A seminal study of this area by Sachs (1967) appears to have also set the methodological and theoretical groundwork for empirical studies of understanding, semantics and pragmatics. From her results, Sachs concluded that the memory for the meaning of a sentence is independent of and more persistent than the memory for the original form of the sentence. Other studies have generally supported these results (Carpenter, 1973; Garrod & Trabasso, 1973; Franks & Bransford, 1974; Schwartz & Witherspoon, 1974; Tzeng, 1975; Kolers, 1975; Layton & Simpson, 1975; Posnansky & Rayner, 1976; Anderson & Paulson, 1977). However, some of these studies tend to support the notion that syntactic factors also affect the processing and retrieval of information (Bock & Brewer, 1974; Bates, Masling & Kintsch, 1978) and that syntactic information is available even in long-term memory (Carpenter, 1973; Garrod & Trabasso, 1973).

Although TG theory can no longer be used to explain such phenomena adequately, there is no reason to discontinue studies of the relation between grammatical complexity and comprehension. As we point out later in this paper, semantic and pragmatic factors cannot explain all comprehension problems (Delis & Slater, 1977); there are no doubt grammatical correlates to such features as propositional density (Kintsch & Keenan, 1973; Kintsch & Vipond, 1977) and topicalization, among other things. Grammatical complexity should continue to be studied, particularly in light of the document designer's need to make complex legal and bureaucratic language more comprehensible to the lay person.

## Semantics

In recent years, the focus of psycholinguistic experiments has been on aspects of meaning, rather than structure. Thus, most of the recent studies of negation have not treated negatives as a grammatical phenomenon so much as a semantic phenomenon. Studies by Just and Carpenter (1971), Clark and Chase (1972), Just and Clark (1973), and Howard (1975), among others, have examined processing of negatives not as a transformational linguistic process, but as part of a larger conceptual semantic process. Among the findings was that inherent negatives such as "unhappy" and covert inherent negatives such as "absent" and "forget" have the same negative effects on sentence processing as overt negatives (Clark & Chase, 1972).

Sherman's (1976) study of sentences with multiple negatives demonstrated that when sentences contained two or more negatives, comprehension suffered a large decrement. This result was interpreted in terms of cognitive overload, as opposed to syntactic complexity. Similarly, Baird (1974) and Anderson and Reder (1974) looked at the cognitive processes involved in verifying and falsifying true and false negative and positive statements, and intepreted their results in terms of semantic memory and matching of congruent information.

Semantics has been found to be a critical factor in other types of studies, as well. Perfetti (1969 a) measured semantic complexity in the surface form of a sentence in terms of what he calls lexical density--the ratio of content words (nouns, verbs, adjectives) to function words (grammatically predictable articles, conjunctions, prepositions). In determining recall results, this measure outweighed other variables such as sentence length, number of words, and Yngve depth. Rosenberg (1969) used a related measure, semantic integration--the predictability of associations among the words in a sentence--and found that well-integrated sentences are better recalled than poorly integrated ones.

Semantic constraints are another factor that can interact with or override syntax in affecting ease of understanding. Natural language use restricts the concepts sentences can express. Slobin (1966) found that passives are as easy as actives in a sentence-picture verification task (used by Gough, 1965, 1966) when the passive expresses a non-reversible idea (e.g., "The leaves are raked by the boy." One would not say "The boy was raked by the leaves.") Herriot (1969) found the same result for concepts limited by practical rather than strictly semantic expectations ("The patient is examined by the doctor."). Other studies of active and passive sentences have focused on the semantic aspects of topic and comment, or thematization/topicalization (Perfetti & Goldman, 1975). And studies of other types of topicalization have served to highlight the semantic, rather than syntactic, aspects of sentence processing (Grieve, 1973; Suci & Hamacher, 1972). Stolz (1967) and Schlesinger (1968) found that the difficulty of the nested (center-embedded) sentence is wiped out when there are semantic constraints on pairing subjects and verbs. In light of such findings, Bever (1968, 1970) suggested that the sentence reader/hearer uses a range of comprehension strategies and bypasses syntax when semantic cues are sufficient for getting the meaning.

Studies of ambiguous sentences have provided a great deal of information regarding the semantic processing of sentences. Studies by MacKay (1966), MacKay and Bever (1967), and Foss and Jenkins (1973) have shown that ambiguous sentences take longer to process than unambiguous ones, even if subjects are unaware of the ambiguity. Other studies have shown that when context is available to disambiguate the meaning, comprehension time is reduced (Foss, Bever & Silver, 1968; Bever, Garrett & Hurtig, 1973). These and other studies also show that even with disambiguating information, all meanings of the sentence are processed, although only one is retained (Foss & Jenkins, 1973; Holmes, Arwas & Garrett, 1977).

Ambiguity that is produced by the use of quantifiers (e.g., "*Every* medicine cures *some* disease") has been the subject of numerous psycholinguistic studies (Johnson-Laird, 1969; Glass & Holyoak, 1974; Rips, 1975; Catlin & Micham, 1975). Much of this research has sought to show why such sentences are interpreted differently in the passive than in the active voice and to provide a semantic theory to explain how people verify sentences. This research is interesting from the point of view of document design, because it relates directly to the types of features that make sentences easier or harder to understand.

Some of the most interesting research into semantics has dealt with how subjects integrate information presented in several sentences or clauses. Some of this research has had to do with anaphoric relations between sentences or clauses (Garrod & Sanford, 1977; Clark & Sengul, 1979). Other studies have involved the processing of incongruent sentences (Hoosain, 1974) or integrating incorrect information that was introduced through repeated statements, leading questions and pictures (Zillman, 1972; Hasher, Goldstein & Toppino, 1977; Loftus, 1975; Loftus, Miller & Burns, 1978; Dooling & Christiaansen, 1977).

## CONTEXT AND DISCOURSE STRUCTURE

Until very recently, psycholinguistic studies of comprehension focused on words, phrases and sentences. Anything larger than a sentence (i.e., discourse) was felt to be too complex to deal with in an experimental context. In the last few years, however, with the growing interest in semantics, pragmatics and cognitive processes involved in inferencing and conversational rules, there have been numerous investigations of various aspects of discourse. Many of these studies relate to the context of discourse--either verbal or pragmatic (real-world)--and provide interesting insights into how people comprehend what they hear or read.

The theory underlying much of the research that we will discuss below involves discourse conventions--conversational rules and rules of speaker/listener interaction which theoretical linguists and philosophers of language have postulated (e.g., Quine, 1964; Chafe, 1970; Freedle & Carroll, 1972; Grice, 1975; Searle, 1975). Conversational rules include rules for presupposition, inference, and implication. Haviland and Clark (1974) and Clark and Haviland (1975) describe a conversational rule between speaker and listener which they call the "Given-New Contract." This is a semantic and syntactic distinction that the speaker (writer) makes--and the listener (reader) unconsciously notes--between (1) information that the speaker believes the listener already knows and accepts as true ("Given") and (2) information that the speaker believes the listener does not yet know ("New").

In light of this theory, several interesting experiments were conducted to study the phenomenon of "idea acquisition and retention" (Bransford & Franks, 1971; Bransford, Barclay & Franks, 1972; Potts, 1977; Hupet & Le Bouedec, 1977; Hayes-Roth & Thorndyke, 1979). Among the findings of these studies is that when subjects are presented semantically-related sentences, they integrate the information into holistic ideas.

When asked to recognize the exact sentences that they heard, subjects often wrongly recognized sentences that contained some or all of the concepts in the original sentences. In effect, subjects integrate "given" and "new" information by creating knowledge frameworks from which they can make certain types of presuppositions (Karttunen, 1974) and inferences (Clark, 1977).

### Presuppositions and Inferences

A presupposition is an assumption that a sentence contains regarding the listener's prior knowledge. It may be signalled by a simple syntactic device such as a definite article ("the"), which presupposes that the reader is familiar with the noun phrase that follows it. Or it may be produced by a complex set of syntactic and semantic factors, as in the famous question "When did you stop beating your wife?" This question *presupposes* that you have a wife, and that you stopped beating her; it *implies* that you did, in fact, beat her, and regularly.

Some studies of presupposition support the notion that subjects treat presuppositions as "given" information, and integrate it into their existing knowledge structures (Offir, 1973; Hornby, 1974; Bock, 1977). Just and Clark (1973) went one step further, and looked at how subjects draw inferences from the presuppositions and implications of affirmative and negative sentences.

Results from other research into inference and pragmatic implication indicate that subjects make inferences based on given information, and then integrate the inference into their memory and knowledge structures (Harris, 1974, 1975; Brewer & Lichtenstein, 1975; Wertsch, 1975; Lesgold, Roth & Curtis, 1979). Such memory structures can apparently be either verbal or imagistic (Potter, Valian, & Faulconer, 1977). There is research which shows that visual information can create inferences in verbal memory (Loftus, Miller & Burns, 1978), and that verbal information may not always be incorporated into an interrelating semantic system but is sometimes encoded into episodic memory, which cannot be used to generate new knowledge (Tulving, 1972; Moeser, 1976).

Bransford and Johnson's (1972) series of experiments was probably the first attempt to demonstrate that providing a context can make previously meaningless discourse meaningful and easier to recall. Other studies support Bransford and Johnson's findings, and show, in addition, that titles can alter the comprehension of a text by affecting which information is selected from the text and how this information is organized in memory (Kozminsky, 1977); that subjects' own backgrounds and mind-sets create contexts (schemata) which provide the interpretative framework for comprehending discourse (Anderson, Reynolds, Schallert & Goetz, 1977; Tenenbaum, 1977); that priming material that is inconsistent with the target passage produces better recall of the passage than a consistent context (Wilkes & Alred, 1978); and that graphic and concrete contexts facilitate recall (Sherman, 1976). All these context effects may be due to the fact that a context causes the listener/reader to make inferences, which provide cues for remembering and retrieving the information in the discourse (Frederiksen, 1975; Masson, 1979).

Other studies have investigated other aspects of context. Tyler and Marsen-Wilson (1977) demonstrated that semantic context influences syntactic decisions in understanding ambiguous sentences. Newstead (1976) showed that a knowledge of the world (pragmatic context) can also influence sentence perception. Pepper and Prytulak (1974) illustrated that context will influence subjects' perceptions of quantitative expressions (sometimes, often, etc.). Underwood (1977) found that in a dichotic listening task, unattended messages served as a context that facilitated understanding of the attended message.

Other recent studies of discourse have dealt with the structure and organization of discourse, and their effect on comprehension (Bower, cited in Wilkes & Alred, 1978; McKoon, 1977; Kintsch & Vipond, 1977; Kintsch & Kozminsky, 1977; Fuque & Phye, 1978). In general, the structure of discourse does affect the way it is understood and remembered. For example, findings from a recent study by Kieras (1979 a) support the existence of a linguistic convention requiring that the theme of a passage be stated first in the passage. Furthermore, information is organized hierarchically in memory, and so subjects understand, remember, and verify topic information better than detail information. Thus, Kieras (1979 b) explains the nature of the overall coherence in technical passages in terms of the reader's constructing a *macrostructure* for the passage, and selecting the central referent of the macrostructure for the statement of the topic.

Miller and Kintsch's (1980) results support and expand this general theme. They point out that "the readability of a text is determined by the ways that certain text properties--primarily the arrangement of the propositions in the text base, but also word frequency and sentence length--interact with the reader's processing strategies and resources" (p. 26). They also point out that two other important determinants of readability are the number of inferences that the reader needs to make in constructing a coherent text base, and the number of times already processed propositions must be reinstated into working memory.

Even in these very cognitively-oriented theories of discourse processing, however, the role of sentence structure cannot and should not be forgotten. Miller and Kintsch's innocent reference to "the arrangement of the propositions in the text base" covers a multitude of syntactic structures: arrangement of propositions is not random, nor is it simply linear. Correct or incorrect focus in a passage can be a result of something as simple as the use of active or passive voice, of coordinate or subordinate clauses. In examining the structure of discourse, psycholinguists cannot afford to neglect grammatical structure. For although it does not answer many of the questions of how we understand written and spoken language, we have a great deal of evidence that many people cannot get beyond the complex syntactic structures of many documents to even begin to process them as discourse.

In summary, we have looked at research into various aspects of written and spoken language, from the point of view of language structures--words, sentences and discourse. We used this categorization not only for our convenience, but also because these are the categories that document designers must deal with. Their needs are concrete, and the research which we have cited relates, to some degree, to these needs. In short, we have tried to review the psycholinguistic literature with the document designer in mind, and to provide a link between research and real-world document design problems.

# II. Cognitive Psychology

Cognitive psychology attempts to explain human thought and behavior in terms of underlying "mental structures," like images, concepts, propositions, plans, or procedures. Where behaviorist psychology emphasizes the importance of observable behavior and learning (see the Instructional Research chapter for the flavor of the behaviorial approach), cognitive psychology emphasizes the importance of internal processes and knowledge. The cognitivist approach is that, in order to understand any mental activity, we must first determine what knowledge people bring to bear on that activity. This approach applies even to the activity of learning, since new concepts can be acquired only with the aid of strategies the learner already knows, and understood only in terms of concepts the learner already has. Bransford (1979) examines this issue in great detail, and we will return to it later in this review.

The cognitive viewpoint is especially relevant to document design, since it provides a framework for studying the readers and writers of documents. In order to produce guidelines that will result in useful, effective documents, the psychologist must have some idea of what readers need to know, what they already know, and how they are likely to apply their knowledge to the task of reading. In order to produce guidelines that document designers and writers can use, the psychologist must have some idea of what designer/writers need to know, what they already know, and how they are likely to apply their knowledge to the task of creating a document.

To understand the cognitive viewpoint more clearly, we begin with a simplified version of it, like that presented by Neisser (1976). Briefly, its main points--widely accepted in the field--are as follows:

- human memory is made up of patterns of information called *schemas*;

- schemas may contain perceptual information--how things look, sound, feel, etc.;
- schemas may contain relational information--how things (and schemas) are related to one another;
- schemas may contain procedural information--how to deal with both external objects and internal schemas;
- a schema may contain all three types of information (e.g., like a set of directions: "Turn left at the first red light, then go three blocks. . .");
- all conscious activities are directed by schemas;
- when an activity yields information missing from the schema (e.g., "What color is it?") or contradicted by it (e.g., "That road isn't supposed to be here!"), the new information is added to memory;
- schemas may be combined to form larger schemas.

More formal treatment of the cognitive perspective can be found in Lindsay and Norman (1972), and Rumelhart (1977), both of which give detailed treatments of perception and text comprehension. Bransford (1979) provides an excellent introduction to comprehension, and Simon (1979) gives a good, if brief, summary of cognitive psychology as a whole.

Not surprisingly, a rich "vocabulary" of schemas seems to be an important part of many kinds of expertise. In chess, for example, expertise seems to rest on the ability to recognize groups of chess pieces as meaningful patterns, and chess players need at least ten years of practice to learn enough patterns to attain mastery (Chase & Simon, 1973). Composers and painters also require a minimum of ten years before they can produce a first masterpiece (Hayes, 1980). Training in law and medicine rests mainly on years of exposure to many thousands of special cases. Others point out (Winston, 1977; McCorduck, 1979) that computer programs that are designed to diagnose medical problems (like MYCIN and INTERNIST), or to identify chemical structures (DENDRAL), or to solve calculus problems (MACSYMA), or to resolve genetic structures, all require large amounts of specialized information. Proficiency in accounting (Bhaskar, 1978), physics (Larkin et al., in press), and algebra-word problems (Hinsley et al., 1977) also seems to be schema-driven, though in these domains the expert tries to apply a small number of problem types to a large number of problems. And, of course, skill in reading and writing demands the recognition of patterns of letters as words, patterns of words as propositions, and so forth.

Because language is one of the most complicated of all human capabilities, its users must acquire a huge body of knowledge in order to master it. Hayes (1978) points out that even the casual reading of a piece of text is often an ill-defined problem. That is, the readers have no way of being certain that they have correctly understood what was read, since they have nothing to check their interpretation against except previous knowledge. The situation of writers is even worse, for they must carefully use language appropriate not only to the specific content area, but also to the intended audience, about whom very little is known.

We already have noted that cognitive psychology is broadly concerned with mental structures, with things that "go on in the head." In the paragraphs that follow, we will attempt to identify and examine some of the properties that are associated with the mental structures necessary for creating and understanding documents. We chose the five properties listed because any discussion of cognitive theories sooner or later touches upon them:

1. *Variety*--the reader/writer must command a wide variety of forms, such as words, clauses, sentences, figures of speech, modes of argument, genres, etc.
2. *Flexibility*--forms which are too specific tend to function as cliches and stereotypes, preventing their user from dealing effectively with new situations.
3. *Hierarchical structure*--lower-level skills, like spelling, must be either mastered or compensated for, to prevent them from interfering with higher-level activities.

4. *Use of non-linguistic information*--textual information can be powerfully supplemented by visual aids, as well as by previously acquired real-world knowledge.
5. *Use of procedural information*--since reading and writing are *activities*, much of the knowledge involved in their performance is about what to do and when to do it; this section will therefore present a detailed model of the writing process, and discuss some of its implications.

## 1. VARIETY

As we have already noted, an important part of expertise is the ability to recognize complex patterns; the experts must have many schemas--and many kinds of schema--at their disposal. Thus the writer who knew only words would be unable to generate sentences; the writer whose skill was confined to sentences would be unable to write good paragraphs; and so on. Interestingly enough, rhetoric, one of the classical bases of language instruction, has already identified many kinds of forms that writers may use as schemas. Although few good writers--or even English teachers--can actually name the various figures of speech or modes of argument, many examples of these structures are easily recognized as cliches or stereotypes (see Lanham, 1968), suggesting that good writers do indeed have mental structures corresponding to these forms.

Psychological research in this area is only beginning. The traditional focus of psycholinguistics has been on the sentence, with grammatical structures and propositions ("kernel representations") as the primary concern. These structures, however, fail to capture much of the richness of language. Below the level of the sentence, grammatical structures fail to explain the effectiveness of figures of speech; above the level of the sentence, grammatical structures fail to explain the use of modes of argument or genres of discourse--more generally, they fail to explain the relationship between the sentence and the context in which it appears. Indeed, since many sentences cannot be understood without a surrounding context--verbal, physical, or social--students of language have been compelled to examine the larger schemas of semantics, pragmatics, and macrostructures of discourse (see our review of psycholinguistics, as well as Section 4, below, for a more complete discussion of these issues).

## 2. FLEXIBILITY

Complex structures like story grammars (Lakoff, 1972) or scripts (Schank & Abelson, 1977) provide a powerful explanation of complex human behavior. The user of a "restaurant script," for example, knows exactly how to behave in a restaurant--how to enter, be seated, order, eat, tip, pay the bill, and leave. Unfortunately, such scripts tend to be so specific that they cannot deal with the world in a flexible way. Thus the user of the restaurant script might be completely baffled by minor departures from it, like a cafeteria-style arrangement, or a salad bar, or a Gypsy violinist. Although it is true that some people show such rigidity most of the time and most people show it some of the time, such behavior is hardly desirable. People must be flexible in dealing with the real world, and psychologists must be flexible in describing their behavior. Hayes' (1978) review of the literature on set and functional fixedness, and Lanham's (1977) work on the use of cliches in college writing are examples of studies in this area.

One possible source of such flexibility is the use of *goals* (Simon, 1969; Schank & Abelson, 1977). A goal is essentially a description of a desired state of affairs without information about how to achieve it. Thus the goal "I must get to the airport (somehow)" leaves the user free to look for alternatives when his usual mode of transportation fails. Furthermore, goals can be combined into much more complex structures called *plans*; for example, "If I want to fly to New York (superordinate goal), I should first make reservations (subgoal), then buy a ticket (subgoal), then get to the airport (subgoal), then . . .".

Such compound structures allow both power and flexibility in dealing with complex, ill-defined problems--like the handling of documents (we will discuss these issues more fully in Section 5).

Formal grammars, on the other hand, are far too rigid to describe the behavior of real users of language. Although human beings can often handle slightly ungrammatical sentences with ease, grammatical theories are very poor at explaining how they do so. Much of the difficulty lies in the assumption that a surface structure--i.e., a sentence--must be completely transformed into a deep structure--i.e., a schema--before understanding can begin. From recent work in psycholinguistics (Cole & Jakimik, 1980), it seems clear that understanding is actually a part of the listening process, aiding the interpretation of an utterance from the very first word. The listener forms hypotheses about each word as he encounters it, without waiting for complete information to become available. Recent work suggests that such behavior is typical of both readers and writers (Hayes & Flower, 1980a).

## 3. HIERARCHICAL STRUCTURE

As Simon (1969) points out, hierarchies are as much a feature of the mental universe as they are of the physical one; as molecules are made up of atoms, so goals are made up of subgoals, and skills of subskills. In writing, the nature of these hierarchies is dictated in large part by the nature of the written product: texts are made up of paragraphs, paragraphs of sentences, sentences of words, and words of letters. Thus, the skilled writer must be fluent in grammar, vocabulary, spelling, and handwriting, printing, or typing. A failure to master one of these skills means more than an inadequate final product; it means that the writer may be forced to plod through a relatively low-level activity at the expense of the rest of his or her writing. In this regard, Hayes and Flower (1980a) found that poor writers often spend much of their attention on low-level activities like spelling, and very little on high-level activities like planning and organization. This result is especially disturbing since Shaughnessy (1977) found that low-level errors dominate the writing of many adults.

One method of dealing with low-level skills is to overlearn them until performance becomes automatic and unconscious (Simon, 1979; see also La Berge and Samuels (1974) for a treatment of the automation of reading processes, and Card et al. (1977) on the skill of text-editing by computer). For writing, perhaps the most ambitious effort in this direction has been the use of sentence-combining exercises, based on transformational grammar, to increase grammatical fluency and sophistication. O'Hare (1973) found that seventh graders trained by this method were indeed able to write sentences as complex as those of twelfth graders, though the effect of this training on other writing skills was not reported.

Low-level skills can also be supplemented by external aids. Using a typewriter, for example, frees children from the mechanics of printing and handwriting (Sandberg, 1979). Similarly, computer programs that identify misspelled words can free adults from the mechanics of spelling, much as text-editing computers (word-processors) can free them from the mechanics of neatness and formatting (Card, 1978). Even when the need for low-level skills has been bypassed, however, the lack of such skills may have previously interfered with *learning* so much that the writer still needs high-level training. This technique of bypassing low-level skills has been used successfully to teach musical composition: a computer helps the student compose not only by playing the music, but also by providing simple musical phrases to work with (Bamberger, 1972). Similar efforts in teaching writing would be interesting, but have not yet been tried. Indeed, the lack of research in this area is most unfortunate, since we badly need information on the relative importance of low- and high-level skills to tell us how they should be taught.

## 4. USE OF NON-LINGUISTIC INFORMATION

Text comprehension requires more than a knowledge of the structure, or even the social conventions, of language. To fully understand a text, we must have some prior knowledge of the reality to which it refers. For this reason, the study of language has begun to place more and more emphasis on semantics and pragmatics (see our review of psycholinguistics).

Similarly, computer models of language processing rely heavily on simplified models of the real world. Such model worlds do more than test a computer program's understanding by giving it something to talk about; they allow the program to clear up ambiguities in a text by referring to an external "reality." Thus Winograd's (1972) program SHRDLU simulated a robot moving children's blocks around on a table top, an environment it understood thoroughly; and HEARSAY, a sophisticated speech recognizer, began with an understanding of chess (Reddy et al., 1973). More general models of the world have also been used: Schank and Abelson's (1977) PAM and TALESPIN understand and generate simple stories by assuming that all the characters have a limited number of goals (e.g., food) and a limited number of basic procedures--"operators"--for satisfying those goals (e.g., find, ask, bargain, threaten, steal); a much simpler program, FRUMP, produces abstracts of news stories about accidents, taking advantage of these stories' standardized content and style. And a more sophisticated version of this approach is used by Hayes and Simon's (1976) UNDERSTAND to interpret texts in which problems of a very common type--involving the transfer of objects from one place to another--are stated. These programs are discussed in detail by McCorduck (1979) and Winston (1977).

For document design, the moral of all these models is that designers must have an accurate picture of their audience; they must find out how much the intended audience already knows about the document's subject matter. Although considerable work has been done on audience analysis, both in rhetoric (e.g., Clevinger, 1966) and psychology (e.g., Karlins & Abelson, 1970), nearly all of this work has concentrated on persuasion rather than content. Needless to say, given the diversity of possible subject matter it is hardly surprising that investigators would give up looking for general underlying principles of content, and focus instead on areas where general principles are more likely to be found. One promising area, which we will discuss in Section 5, replaces the study of the audience with the study of the process by which a writer analyzes that audience.

Often, document designers must provide some background to enhance understanding by their audience. One way to do this is by means of non-textual forms of information, like pictures (Bransford & Johnson, 1973). But even when the audience has enough background to understand the subject matter, it is often better to present information in some form other than prose. Pictures, for example, are often easier to remember than verbal descriptions (Gehring et al., 1976). Furthermore, representations like pictures, drawings, graphs, flowcharts, and tables are often better than text at making certain kinds of relationships clear. Thus Wason (1968) found that logic trees transmit conditional information--exceptions, qualifications, etc.--better than prose. Wright (1971; Wright & Reid, 1973) used different representations of complex relationships to show that the appropriateness of a representation depends on both the nature and the difficulty of the problem it is being used to solve. Wright found that tables are the quickest for solving easy problems, if the user knows what to look for; flowcharts are the most accurate, especially if the user does not know what to look for, or if the problem is hard; lists of short sentences are best if the material is to be memorized; and bureaucratic prose yields the slowest and least accurate problem-solving, but performance improves if the user is forced to work from memory (e.g., when a set of instructions precedes a large number of short problems). And finally, even if special representations are not used, text can still be made more effective by the careful use of headings (Kominsky, 1977) and special formats (Dever & Friend, 1976). These and related issues are also discussed in the human factors and typography chapters of this review.

## 5. USE OF PROCEDURAL INFORMATION

Because reading and writing are activities, much of the knowledge used in their performance is procedural knowledge, that is, knowledge about what to do and when to do it. Since we have already treated the reading process at length in our review of psycholinguistics (see also Section 4, above, for computer models of comprehension), we will not discuss it further here. Instead, we will examine the process of writing in some detail. As a starting point, we will take the model of Hayes and Flower (1979). There are at least three reasons for this choice: first, the features of the model are relatively straightforward; second, the model is complex enough to cover a wide variety of writing behavior, and precise enough to be experimentally tested; and third, the model is backed by an enormous amount of data, over 100 hours of carefully analyzed tape-recordings of writers at work. The main features of this model are as follows:

Writing is goal-directed. Hayes and Flower (1980b) identify three kinds of goals: rhetorical goals (e.g., "I've got to get their attention."), product goals (e.g., "This has to be short."), and procedural goals (generating ideas, organizing ideas, translating ideas into text, and editing the text). Note by the way, that all goals must be implemented by procedures, and therefore require procedural goals. A rhetorical goal, for example, can involve idea generation ("What do they want to know?"), organization ("What will they need to know first?"), translation

("How can I make this clear?"), and editing ("Will they understand this?").

Writing is hierarchically organized. More precisely, goals may have subgoals, and subgoals may have other subgoals, giving the writing process a hierarchical structure (see Sections 2 and 3 above). Thus if we want to write a report, we might first make an outline; to make an outline, we might first generate ideas and then organize them; to organize the ideas, we might look for likely topic headings; and so on.

Writing is text-driven. If writing was a purely hierarchical process, it would proceed much more systematically than it does. In fact, writers frequently interrupt their work to make changes in spelling, punctuation, word choice, grammar, and even to check product goals ("This is too long.") and rhetorical goals ("They'll never buy that!"). Much of this behavior seems to be caused simply by noticing errors or getting new ideas while rereading the text that has just been written.

Writing processes are often recursive. In other words, in order to achieve a goal, sometimes we must satisfy an even higher-level goal (the definition of "recursive" is actually a bit more complex than this, but we need not deal with it here). For example, if we were *editing* a sentence (low-level goal) and discovered that some of the words in it needed to be explained, we might then be forced to *write* a whole paragraph (high-level goal) to supply the needed explanation.

Goals may be modified. Many writers, particularly experts, will modify or abandon a goal if they feel that they cannot satisfy it. Thus a writer might eventually abandon a topic as being too technical, or controversial, or uninteresting; or the writer might decide to treat the issue as part of something else. One possible explanation for this goal-modifying behavior is that expert writers may treat at least some of their goals just like text, so that these goals can be generated, organized, translated, and edited just like any other kind of text. Some of these features are foreshadowed in the work of Emig (1971), whose study of 12th grade writers revealed such processes as "planning," "reformulating," and "stopping and contemplating product."

Discussion. From the work of Flower and Hayes (1980) it appears that an important part of writing expertise is an ability to take full advantage of these features. Thus experts tend to formulate complex goals and change these goals in the course of writing, while novices are more likely to stick to very simple plans--generate an idea, translate it into a sentence, generate another idea. Experts often reread their work while they are writing, both to correct the work and to generate new ideas; novices, on the other hand, are much more likely to ignore what they have written, or only check it for minor errors. And, most important for document design, experts are far more likely to formulate rhetorical goals, such as paying attention to the needs of their audience.

Unfortunately, the novice's lack of attention to his audience is most likely to occur in exactly those situations in which documents are normally written. Atlas (1980) found that novice writers who were asked to reply to a business letter usually ignored the letter writer's concerns, even when they understood these concerns quite well. Novices concentrated instead on the outline they had been given as a writing aid. Since most documents are written according to guidelines, and with no direct contact with the audience, bad document writing may well have similar causes. The fact that Atlas' writers improved greatly when the outline was removed, and improved even more when the audience's needs were made more explicit, suggests that document designers might well benefit from similar procedures. Indeed, Scardamalia, Bereiter, and McDonald (1977) found that even the writing of elementary school students could be improved by audience sensitivity training (by videotape). This suggests that contact with the intended audience is an important part of the design process.

So far we have discussed how to improve the written product, but how do we improve the writer? Flower and Hayes (1977) suggest that many of the expert's high-level planning strategies are well-known in the psychology of problem-solving, and can be explicitly taught. Whether a mastery of these techniques, together with the low-level skills, discussed in Section 3 above, is enough to make a writer an expert, we do not yet know.

# III. Instructional Research

Education, when viewed as a domain of research and practice , shares several fundamental concerns with the emerging field of document design. For one, education and document design both are concerned with the same essential learning processes that govern what humans comprehend and retain from written materials. Second, education's interest in the relationship between the content to be learned and the learner is analogous to the relationship between a document and its user. And third, both education and document design share the need to reach a vast cross section of people who vary in individual abilities and experiences.

The particular educational research subarea that provides the most powerful insights for document design issues is instructional research. This body of research focuses on the factors operating in the instructional process--the instructional materials, the nature of the learning task, and the characteristics of the learners.

A very basic distinction must be made between traditional learning research and instructional research. As articulated by Bruner (1964) and Glaser (1976), learning theory and research is *descriptive* in the sense it attempts to describe, after the fact, the conditions under which learning occurs. In contrast, instructional theory and research is *prescriptive* in the sense that it attempts to establish rules that specify the most effective way to attain knowledge or master skills. Instructional research centers on the development of procedures for optimizing learning in specific situations. Because document designers are tasked with developing documents for specific audiences, situations, and purposes, instructional research clearly is an appropriate body of knowledge in which to seek possible answers and guidance.

Interest in instructional research has been accelerating since the 1960s. An indicator of this interest is reflected in the space given to instructional psychology in recent issues of the *Annual Review of Psychology* (Gagné & Rohwer, 1969; Glaser & Resnick, 1972; McKeachie, 1974; Wittrock & Lumsdaine, 1977). The research runs the gamut from theoretical, laboratory studies to field evaluations of specific curricula; it has examined all ages of learner groups and many different content areas.

The instructional research literature itself has become too large and diverse to be subsumed neatly as a subcategory under educational research. Indeed, many studies which can be classified as "instructional" come from cognitive psychology, verbal learning research, psycholinguistics, and human factors--all areas that, strictly defined, are outside the domain of educational research.

Our purpose in this review is to examine the instructional research that reasonably can guide the development of effective documents and help pinpoint the gaps in our knowledge. We have restricted this review, therefore, to four instructional research areas that are directly relatable to document design issues and practices. We feel that these research areas most likely will yield information of immediate practical value. The areas and our rationale for selecting them are:

> *Programmed Instruction.* This research is relevant in that PI consists mainly of meaningful written prose. PI parallels document design in the sense that prose content is presented in units ranging from single sentences to several pages. PI has been offered in a variety of formats that can be duplicated in other kinds of documents.
>
> *Prose Learning.* This research is relevant in that prose is the kind of writing contained in most public documents. The psychological processes that affect comprehension of instructional prose presumably are similar to those for understanding prose in public documents.
>
> *Aptitude-Treatment Interaction.* This research is concerned with the fact that different individuals learn better from different instructional treatments.

> It is relevant in the sense that it seeks ways of optimizing instruction for different kinds of learners.
>
> *Instructional Technology.* Instructional technology is not a research discipline. The technology is derived from research, however, and is relevant in the sense that it offers empirical procedures for systematically developing and testing documents.

We recognize that, in the absence of a specific document design research literature, these instructional research areas can only approximate the conditions involved in the design and use of real world public documents. This obviously restricts the generalizations one can make to practical applications. Nevertheless, both educators and document design practitioners are fundamentally concerned with meaningful verbal learning and the methods for facilitating this learning. This certainly justifies looking at instructional research as a beginning point for developing a document design literature.

## PROGRAMMED INSTRUCTION

The programmed instruction (PI) movement initially emerged by efforts to apply operant learning principles derived from the laboratory to instructional settings. PI is characterized by presenting relatively small instructional segments, allowing learners to make responses during the instruction, and verifying the accuracy of responses before proceeding to the next segment. The earliest research compared PI against other instructional modes and compared the relative effectiveness of different styles of PI. A review of this work (Schramm, 1964) found mixed effects but pointed out the relative futility of compared effectiveness studies. The need to examine instructionally relevant variables became a more pressing research issue. Later PI research thus began to examine program variables presumed to affect learning. Illustrative examples of this research follow.

### Knowledge of Correct Response

Does learning increase when learners are provided answers to check their responses while reading? A review of the research in this area found that providing immediate knowledge of results at frequent intervals enhances learning (Schramm, 1964), although contradictory findings have been reported (e.g., Glaser & Taber, 1961). Krumboltz and Bonawitz (1962) found providing feedback in the context of a sentence was superior to presenting the feedback term alone. Still other research (Brackbill, Wagner & Wilson, 1964; Briggs & Hamilton, 1964) found that delaying feedback during learning improved retention over a wide range of tasks ranging in meaningfulness and difficulty.

### Prompting

Does giving learners clues about important content before they respond (as opposed to confirming the correctness of their responses) result in higher achievement? Levine (1965) found that prompting in a paired-associate learning task was better than confirmation regardless of degree of meaningfulness. Stolurow and Lippert (1964) found that prompting in teaching educatable mentally handicapped children a sight-vocabulary resulted in fewer training errors than confirmation procedures. Jenkins and Bailey (1965) demonstrated that simple typographical prompts such as color and underlining effectively distinguish important from unimportant content, but too many combinations of cues (e.g., color plus underlining plus full caps) is confusing. Anderson and Faust (1966), however, found underlining did not help in learning Russian words. Color or underlining prompts were found more effective in aiding readers to attend to critical portions with complex stimulus displays such as textbooks or long prose passages (Anderson, 1967; Klare, Mabry & Gustafson, 1955). Reviews of research in this area generally conclude that prompts can enhance most kinds of learning by making important content more salient (Anderson, 1967; Briggs, Campeau, Gagné & May, 1966).

## Response Mode

Do covert or overt responses in programmed materials make any difference in learning? In general, research indicates that there is little difference in learning between overt and covert responses, although the active participation due to either response mode is better than straight reading (e.g., Della-Piana, 1962; Crist, 1966). Studies have shown that overt responses are helpful if the response is relevant to critical content to be learned (Holland, 1965), when response learning such as technical language is the requirement (Williams, 1963), and when responses are difficult or novel for the learner to make (Goldbeck & Campbell, 1962; Krumboltz, 1964). Reviews of this research area are presented by Schramm (1964), Glaser (1965), and Anderson (1967).

## Step Size

Is it better to give smaller or larger bits of information before learners are asked to make a response? Although this research is ambiguous because the term "step size" is used differently, evidence favoring small step size was reported by Evans, Glaser, and Homme (1960; 1962) and Coulson and Silberman (1960). Frase (1967) found that 20-line prose passages are optimal for retention with non-programmed instructional material. In research comparing different PI styles, Bisbicos (1965) found no difference in learning between linear programs (smaller step size) and branching programs (larger step size). Pressey and Kinzer (1964) found that a small-step, linear program was no better than reading a text with adjunctive questions, but took far more time.

## Some Learner Variables

From early PI research it became clear that individual differences among learners interact with program variables. Campeau (1965), for example, found a significant interaction between anxiety level and the presence or absence of feedback in PI. Porter (1961) reported that greatest gains in learning spelling were made by low-IQ students taught by PI. McNeil (1962) found that in teaching elementary reading skills by PI, the requirement for oral responding was more effective for low-IQ than for high-IQ children. Reviews of other studies on the interaction between program and learner variables can be found in Briggs, Campeau, Gagné and May (1966) and Anderson (1967).

In assessing the overall value of PI research for document design we can note several contributions. First, PI research stimulated the evolution of an empirically-based technology of instruction that is equally applicable for the design, development, and evaluation of any prose-text document. We will discuss this in more detail in the Instructional Technology section of this review. Second, PI research isolated a number of instructional and learner variables that affect what is learned from written instructional text. These same variables doubtless influence comprehension from other kinds of documents as well. And third, PI research demonstrated the complexity of the instruction-learning process, a complexity which very likely is duplicated by the document-comprehension process.

## PROSE LEARNING

Over the past ten to fifteen years, interest in how learning takes place from written prose materials has increased. The rationale underlying this research is straightforward: since a significant proportion of today's communication consists of written prose, it is a "given" that needs to be better understood. This research is directly linked to document design in that the research largely has been conducted on meaningful prose, the kind of prose typically found in textbooks and other everyday written materials.

Two extensive reviews of the research in this general area have been completed. Carroll (1971) surveyed the research done up to 1970 on the comprehension of meaningful connected discourse. Carroll covered research done on both spoken and written verbal messages presented by a variety of educational media. McConkie (1977) reviewed the more recent research related to learning from written text. Both of these literature reviews, like the present review, encompass research from a number of fields, among them cognitive psychology and psycholinguistics.

The present review examines a fairly narrow slice of the prose research literature. We are concerned specifically with prose research that has been conducted on instructional materials in educational settings, and especially on those prose learning experiments that involved altering features of prose text.

One such type of alteration consists of supplementing the prose text with different forms of adjunct materials to see if they enhance learning. Advance organizers, overviews, and inserted questions are examples of adjunctive materials. A second type of alteration consists of varying the organization, sequence, and wording of sentences that comprise the prose passages. We review the research on both types of alterations in the following sections.

### Advance Organizers

David Ausubel, a psychologist, was among the earliest to recognize the need for prose learning research on the hard-to-refute grounds that most instruction entails meaningful connected discourse, like textbooks and lectures. Classrooms and other instructional settings require reception learning in which content is presented (e.g., as in reading) for the student (or reader) to internalize and make available at some future time.

Ausubel (1963, 1968) developed a cognitively-oriented theory to account for this kind of learning. The theory essentially postulates that learning occurs because meaningful materials are related to concepts already known.

The most important single factor influencing learning in Ausubel's system is what the learner already knows (i.e., the proactive effect from learning theory). Ausubel devised advance organizers as the means for facilitating the integration of new concepts with those existing in cognitive structure. Advance organizers are "introductory materials . . . introduced in advance of learning . . . and presented at a higher level of abstraction, generality, and inclusiveness . . . (in order) to provide ideational scaffolding for . . . retention of the more detailed and differentiated material that follows" (Ausubel, 1968, p. 148). Advance organizers thus function to bridge what individuals already know with what they are in the process of learning. Advance organizers work best when they precede material that is unfamiliar, technical, or otherwise unrelatable to past experience.

Advance organizers seem potentially adaptable to document design in at least two ways. First, advance organizers are short introductory prose segments, written in a special way, that are placed before larger prose passages. Consequently, they theoretically are capable of being inserted quite easily in many different kinds of documents. Second, advance organizers are intended to aid the assimilation of new material to existing knowledge. This certainly is consistent with the intentions of many public documents. The value of advance organizers for document design ultimately will rest on how effectively they enhance learning by facilitating the subsumption of new concepts under ones existing in cognitive structure. The evidence on this facilitating role is mixed.

Some early experiments by Ausubel and others showed promise but later research has not yielded clear support. For example, Ausubel and Fitzgerald (1961, 1962) found that advance organizers increased retention of 2,500 word prose passages better than other forms of introduction, but the effect was most pronounced for the subjects with below average intelligence. Allen (1970) confirmed this finding. Other data (Fitzgerald & Ausubel, 1963) indicate advance organizers significantly increased retention scores for subjects in the upper third of verbal reasoning ability. Ausubel and Youssef (1963) found that advance organizers which preceded a passage on Buddhism resulted in higher retention than a non-organizing historical introduction. Further research support has been provided by Wittrock (1963), who found that advance organizers contain a learning set that facilitates learning, and by Scandura and Wells (1967), who found advance organizers in the form of a mathematical game facilitate learning abstract mathematics.

Other research has failed to support the facilitative effects of advance organizers. Jerrolds (1968) compared ninth-grade students on advance organizers, more typical overviews, and no organizers to determine learning effectiveness. No differences in learning were found. Davis (1969) found no differences between subjects who memorized organizers before studying prose passages and those memorizing organizers after. Kirkwood (1971) did not find any differences in learning about mass production among groups given advance organizers, motivational passages, or placebo introductions. Barron (1971) tested groups of sixth to twelfth grade students on general science content with prose advance organizers, graphic advance organizers, and no organizers. No significant differences in favor of advance organizers were found at any grade level. Graber, Means, and Johnston (1972) studied the effects of pre- and post-organizers and a historical passage with 96 undergraduate chemistry students. They, too, found no significant differences. In a review of 32 advance organizer studies, Barnes and Clawson (1975) found that only 12 reported significant results. These researchers concluded that "advance organizers, as presently constructed, generally do not facilitate learning" (p. 651).

The apparent lack of clear empirical support for a theory of advance organizers is still being debated. Lawton and

Wanska (1977), in replying to the Barnes and Clawson review, claim that there is much misunderstanding in the research literature about the meaning and purpose of advance organizers. To Lawton and Wanska, the critical feature of advance organizers is that they facilitate learning only when the new material to be learned substantively relates to more inclusive subsuming concepts in cognitive structure. Since these subsumers already may be present from prior knowledge, the lack of differences between advance organizer and control treatments may be attributed to prior knowledge. Since it is not clear if prior knowledge was adequately controlled, the reported findings of no difference do not necessarily disprove the efficacy of advance organizers.

In a defense of his theory and research, Ausubel (1978) points out that many studies on advance organizers are flawed because (1) they do not systematically appraise the relevant concepts already in learners' cognitive structures, and (2) criterion tests that require only verbatim retention of material are used even though advance organizers are designed to favor meaningful learning (e.g., applying concepts to novel problems). Mayer (1978) provided support for the latter point in an experiment with college students learning from a text about computer programming. Mayer found that advance organizers interact with types of learning performance. Advance organizers enhance meaningful learning, such as recalling conceptual ideas and making novel inferences, but not the learning of technical facts.

Another organizational strategy that is akin to advance organizers is overviews. Like advance organizers, overviews are almost always presented immediately prior to the content to be learned. Unlike advance organizers, overviews are familiar to most students and readers and are commonly used in many different kinds of written documents. Overviews also differ from advance organizers in that they are introductory materials presented at the same level of abstraction and inclusiveness as the text material that follows them.

There is no theory to explain the effects of overviews.

Overviews presumably help establish productive learning sets by aiding the reader to grasp the major topics and organization of the main body of information. Little research has been done on overviews, even though they are widely used and have been around a long time. Weiss and Fine (1956) reported that including key words of the content to be learned in an overview improved retention. Reynolds (1966) found that giving students a perceptual organization of factual materials by a map increased learning. In a literature review, Hartley and Davis (1976) concluded the scattered evidence on overviews suggests that they may aid learning, but much more needs to be known about their construction, length, and effects.

### Inserted Questions

Another way to alter the features of prose text by adjunctive materials is to insert questions in existing prose narrative. Whether the questions are located in another document and referred to in the prose text or they are placed within the prose text itself, the semantic meaning of the original prose is not modified. The inserted questions thus are "adjuncts" to the prose narrative.

Education has a long-standing interest in the effects of questions and research can be traced back to the early years of this century. The more recent, productive research on questions stems from Ernest Rothkopf's theoretical and experimental work in written prose learning at Bell Laboratories. This line of research appears especially relevant to document design because it gives insight and perspective on natural language comprehension and human information processing.

Rothkopf (1965, 1970) coined the term mathemagenic behaviors to describe the activities that "give birth to learning" when readers contact written text. Mathemagenic behaviors encompass such reading activities as scanning, segmenting, and processing. The theory asserts that adjunct questions, when interspersed within written prose text, control these reading activities by shaping the reader's attention and processing activities.

The basic research paradigm used to test these assertions consists of placing questions either before (prequestions) or right after (postquestions) prose passages and asking subjects to answer the questions while studying the passages. Posttests are then given. The posttests consist of test items that cover both the same material as the inserted questions and other material not covered by the inserted questions. If groups who receive the inserted questions perform better on posttests than control groups who read the same prose passages but without inserted questions, then researchers can claim that inserted questions effectively shape reading strategies.

The original research was conducted by Rothkopf and his colleagues. Rothkopf (1966) investigated whether adjunct questions preceding or following prose passages had any effect on learning. Both prequestions and postquestions increased learning of question-relevant information, but only postquestions facilitated incidental learning (information not covered in inserted questions). Data supporting the facilitative effects of postquestions were also obtained by Rothkopf and Bisbicos (1967). Frase (1967) varied passage length, question placement, and knowledge of results and found that postquestions facilitated incidental learning and 20-line prose passages were optimal for retention. Frase (1968) further found that inserted postquestions facilitated incidental retention and that retention increased with the frequency of postquestions. The pattern of results that emerged from these studies was that both pre- and postquestions facilitate learning question-relevant information, but only postquestions facilitate incidental learning.

This research demonstrated that the position of inserted questions is an important variable in learning from prose text. Additional variables have been examined in other research. One such is the frequency of inserted questions. Frase (1968) used a 2000-word passage about the life of psychologist William James that was divided into 20 paragraphs of 10 lines each. Frase found better performance when questions were inserted after one paragraph than after five. Boyd (1973) supported this finding with different prose content.

Motivation to answer inserted questions is an example of another variable that has been studied. Frase, Patrick, and Schumer (1970) created three levels of motivation by giving subjects either 0¢, 3¢, or 10¢ for each correct answer on the criterion test. They found that inserted questions aided learning more when motivation was low. Rothkopf and Bloom (1970) gave inserted questions either orally (more motivating) or in writing (less motivating), and discovered the oral group did better on a posttest. They found the opposite effects of the Frase study: motivation can enhance questioning effects.

Another variable that has been given close research scrutiny is the type of inserted question. Inserted questions can be categorized by the nature of cognitive processing required to answer them. At the low end of the level-of-processing continuum are verbatim questions, which require the learner to recognize information precisely as it was presented in the prose text. At the opposite end are "higher order" questions, which require the learner to comprehend relationships and apply concepts and principles to novel situations. These questions involve deeper levels of cognitive processing in order to be answered. Consistent with the cognitive viewpoint that there are various possible levels of processing (Craik & Lockhart, 1972) one would predict that inserted questions that require comprehension of prose text will promote deeper processing, and hence, more learning than verbatim questions.

The issue clearly is as relevant for document designers as for educators. Different public documents are intended to serve different purposes. Some documents attempt to get their readers to follow specific procedures or do things as prescribed (e.g., patient package inserts in prescription drug containers; directions for operating a machine). Other documents expect readers to understand principles and to apply them in their own situation (e.g., tax instructions). Consequently, if inserted questions really do affect criterion performance by influencing the processing activities of readers, then the study of different types of inserted questions seems a sensible thing for document designers to do. We will review several representative experiments.

Brunning (1968) compared statement-type and completion-type review items after prose passages. He found that the completion-type questions resulted in higher learning. Watts and Anderson (1971) used 300 high school seniors to compare postquestions that either were repeated examples from the prose passage or required the selection of new examples different from any described in the passage. Subjects who received the new-example questions performed better overall on the criterion test. Shavelson, Berliner, Ravitch, and Loeding (1974) compared the effects of higher order and lower order pre- and postquestions. Higher order postquestions produced better performance. Felker and Dapra (1975) reported results that generally confirmed these findings. College students were given fifteen 20-line passages on psychological conditioning with either verbatim or comprehension questions before or after each passage. Comprehension postquestions facilitated problem solving performance on the criterion test over any other condition. In a review of the research done on different types of inserted questions, Andre (1979) concluded that, in general, questions that require comprehension and deeper levels of processing facilitate learning and retention.

The research on inserted questions in prose learning has matured to the point that meaningful reviews and syntheses are possible. Anderson and Biddle (1975) provided the first comprehensive review of the general literature on inserted questions, followed soon after by a more restricted review by Faw and Waller (1976). More focused reviews have been done by Rickards (1979), who critically analyzed the methods and processes used in experiments on inserted questions, and by Andre (1979), who reviewed experiments done with different types of questions.

As a general summary of inserted question research, we can make several comments. First, research with different age groups and different prose content suggests that questions influence reading activities and learning. Second, the type of inserted question apparently can affect both the level of cognitive processing and the level of comprehension attained during the reading of meaningful prose material. And third, the explanations and theories being advanced to explain the effects of inserted questions are increasingly cognitively oriented, a shift that brings this line of instructional research closer to more basic learning research (Anderson & Biddle, 1975; Faw & Waller, 1976; Rickards, 1979; Andre, 1979).

For document design, inserted questions would seem to have potential for increasing the comprehension and relation of material contained in public documents. Obviously inserted questions are a somewhat awkward format to apply directly to most public documents. But the consistency of the results reported in the literature suggests that the essential notions underlying inserted questions are powerful and, apparently, valid across a variety of prose content areas and kinds of people. As such it would seem worthwhile to explore further how these essential notions from instructional materials can be approximated and generalized to public documents.

## PROSE TEXT MANIPULATIONS

Advance organizers and inserted questions are examples of adjunct materials that supplement and enhance the prose text that is to be learned. The prose text has its own meaning independently of the adjuncts. We turn now to some of the research that has studied how the prose text itself can be altered to enhance learning. In this section we review several representative studies to illustrate the nature of the research and to connect this research with document design.

English prose can be manipulated in many different ways. The underlying rationale of research in this area is that the sequence, organization, and wording of sentences affect encoding activities and other information processing systems and, thereby, influence learning. For example, Frase (1969) investigated the organization of text sentences in which names and attributes of things were treated as superordinate items and attribute values as subordinate items. Sentences either were grouped randomly, by names, or by attributes. Systematic groupings produced better recall than random arrangements, a finding confirmed by others (e.g., Myers, Pezdek, & Coulson, 1973; Perlmutter & Royer, 1973). In another study, Frase (1970) studied whether adults given text arranged in logical order retrieve more information than those given text not so organized. He found that both groups recalled sentences equally well, but the group that read the organized sentences was able to derive inferences better. A reasonable conclusion to draw from this set of research is that the sequence of verbal information can determine how information is stored in memory and retrieved.

Well-written prose text is expected to help readers build a cognitive representation (i.e., identify concepts and perceive patterns of relations) of information stated in the passage. A number of experiments have examined a variety of textual manipulations to determine which aid , or detract from, the building of coherent representations of content. Wolk (1974), Bruning (1970), and Meyers and Boldrick (1975), among others, found that scrambling or arbitrarily reordering sentences decreases learning. Such reordering is disruptive because it distorts temporal and causal relations and distributes related information farther apart in the passage. Other researchers (Bransford & McCarrell, 1974; Dooling & Mullet, 1973) have found that when referents of nouns and verbs are left ambiguous, readers cannot establish coherent representations.

Prose text can be segmented into different content groupings in ways that facilitate comprehension. For example, Frase and Schwartz (1976) conducted several experiments on the relation between meaningful text segmentation and comprehension. Using sentences that contained several components from the technical material to be learned, they compared standard text format against text that segmented the sentence components by different forms of indentation. Segmented text resulted in 14 to 18 per cent faster response time. The researchers concluded that both segmentation and indention control comprehension, possibly by aiding reading operations on separable groups of clauses.

It is difficult to argue with the statement that the manner in which prose is presented can influence how well it is understood. Our review of the few studies in this section has suggested just several ways that prose text can be manipulated. Some prose text manipulations contribute to clarity and understanding, others hinder; some manipulations aid information processing during reading, others do so to a lesser degree. The task facing document designers--researchers and practitioners alike--is to discover the characteristics of prose text which influence what readers understand and retain. Or said differently, the task is to find out what "clear writing" is and what makes it so.

## APTITUDE-TREATMENT INTERACTION

Educators and researchers alike have longstanding commitments to quality instruction. Educators over the years have experimented with any number of practical strategies to improve the efficiency and effectiveness of learning for students. Open classrooms, team teaching, films and television, and peer tutoring are examples. Researchers have studied a host of instructional and learner variables in hopes of helping to convert research and theory into workable educational applications.

Even a casual observation of schools and the abilities of their students indicates that education often fails. No single instructional process provides optimal learning for all students. Educational alternatives come and go but still many students do not learn. Clearly, research has not yet yielded sufficient guidance to help educators develop effective instruction for all.

Part of the reason for this lies in the historical roots of scientific psychology and the two divergent research paths leading from them (Glaser, 1972). One path, psychometrics, emphasizes the nature of individual differences. The research dwells on methods for measuring these differences. The other path, experimental psychology, concentrates on the discovery and formulation of general laws of behavior, including laws of human learning. Because these two parallel research thrusts were unconnected, educators were left with assumptions about individual differences untempered by any knowledge of learning, and with theories of learning uninfluenced by the effects of individual differences. Traditional research thus failed to help educators devise different instructional practices that meet the specific needs of different types of students.

For research to be responsive to the practical requirements of education, more will have to be known about how to mesh instructional treatments with individual student aptitudes. Cronbach (1957) was among the earliest to articulate the need for experimental psychologists and psychometricians to combine their methods and to conduct research to find specific aptitude-treatment interactions (ATIs). By educators knowing the specific aptitude patterns that benefit from specific instructional approaches, they will be better equipped to adapt instruction to individual student needs. And since document designers are faced with the parallel issue of designing documents that will be effective for specific users, ATI as a concept seems worthwhile to explore in this context.

A number of definitions and explanations of ATI are available (e.g., Cronbach, 1957; Glaser, 1972; Parkhurst, 1975). In general, an aptitude is any personological variable upon which individuals differ (e.g., IQ, achievement motivation, anxiety). A treatment is any instructional strategy (e.g., linear programmed instruction, competency-based instruction, lecture). An aptitude-treatment interaction exists when individuals at one end of an aptitude variable perform best with one type of instructional strategy (treatment) and individuals at the other end of the aptitude variable perform best with a different instructional strategy.

Figure 1 depicts the classic aptitude-treatment interaction, where $T_a$ and $T_b$ refer to alternative instructional treatments. To maximize instructional effects across learners with different levels of "Y", an educator would ensure that $T_a$ is given to students low on aptitude "Y" and $T_b$ to students

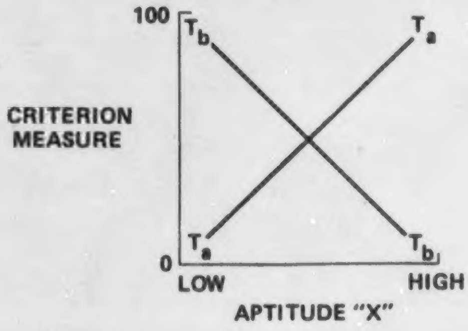

**FIGURE 1. Plot Showing an Aptitude-Treatment Interaction**

high on "Y". Applying this notion to document design, the objective is to design public documents with certain format and language features for readers having certain educational backgrounds and language abilities.

Bracht (1970, p. 627) points out that "the goal of research on ATI is to find significant disordinal interactions between alternative treatments and personological variables." Interaction effects are disordinal, for example, when instructional treatment $T_a$ (cf. Figure 1) is significantly better for low aptitude students but treatment $T_b$ is significantly better for high aptitude ones. Interactions that do not meet these criteria, called ordinal interactions, rarely are sufficient evidence for educators to develop alternative instructional treatments for different student abilities.

Much interactive research has been done in the past, but most of these studies were not specifically designed to find ATIs. What is recognized as ATI research dates from the 1960s. To cite but a few examples, Tallmadge and Shearer (1969) found significant interactions among inductive and deductive reasoning, different subject matters, and individual difference variables, but were unable to replicate these findings in a later study (Tallmadge & Shearer, 1971). Campeau (1965) found that students scoring high on test anxiety learned more from instruction with feedback while low anxiety students learned more without feedback. Skanes, Sullivan, Rowe, and Shannon (1974) studied ATIs between IQ level and ability to transfer information with and without the use of pretests. These investigators found an IQ by pretest interaction.

Several reviews of ATI research have been conducted (e.g., Bracht, 1970; Berliner & Cahen, 1973; Allen, 1975; Tobias, 1976, Cronbach & Snow, 1977) from which we can estimate the potential of the research for educational practice. Bracht (1970) analyzed 90 research studies published in a variety of sources. He found disordinal aptitude-treatment interactions in only five of these studies, which suggests that many aptitude and treatment variables will be of little value for developing effective instructional alternatives. Allen (1975) was able to construct some instructional prescriptions on the basis of a review of research that related "intellectual ability" to media design. The prescriptions are presented as suggestive only, however, with the caveat that the research is too fragmentary and inconclusive to yield valid guidelines for educational practice. In the most comprehensive review done yet, Cronbach and Snow (1977) examined and analyzed numerous experiments involving interactions. They concluded that the existence of ATIs has been clearly established, but few are well understood and none are yet applicable to instructional practice. A large portion of the Cronbach and Snow work consists of suggested remedies that must be applied if ATI research is to yield useful results. Some of the remedies revolve around the need for a better conceptualization and organization of aptitudes and treatments; other remedies focus on statistical and methodological considerations.

What are the implications of this for document design? For one, ATI as a concept unarguably seems relevant. Public documents are intended for all citizens. The array of individual differences represented in any citizen pool is far broader and more complex than even educators face. If we accept the notion that a prime responsibility of document designers is to adapt public documents to meet the specific needs of its users, then ATI clearly is a valid concern.

A second implication centers on the kind of ATI studies that document design researchers ought to undertake. The research done by psychologists and educators in the past suggests that future ATI research must incorporate refined notions of both aptitudes and treatment. Only then will document designers learn much about how to adapt documents to individual user needs. Recent trends indicate that a research approach that views individual differences in terms of cognitive processes is promising. Glaser (1972) has pointed out that we already have identified a variety of cognitive processes involved in new learning and it should be possible to analyze individual differences along these process dimensions. As for the "treatment" side of ATI, much conceptual groundwork will have to be laid before we can derive categories that make sense for document design. We might design a government regulation with either black or red headings, or with two or three columns, but whether these constitute different document "treatments" is problematic.

To summarize, ATI seems to represent a concept and research methodology that usefully can be transferred to study document design issues. Education's experience with ATI has shown some obvious pitfalls to avoid. It is up to document design researchers to build from this experience and extend the application to meaningful document-user interactions.

## INSTRUCTIONAL TECHNOLOGY

Instructional technology is an outgrowth of the longstanding efforts of educational practitioners and researchers to improve classroom instruction. As is true of any other technology, instructional technology (IT) refers to a practical art using scientific knowledge. In this case the art is education and the science is the psychology of human learning. Unlike the other instructional research areas we have reviewed, IT is not a theory-driven field of research per se, although it has evolved from the confluence of several research fields.

IT has been recognized as an entity since about the early 1960s. The most widely accepted definition is the one provided by the Commission on Instructional Technology formed by Congress:

> Instructional technology . . . is a systematic way of designing, carrying out, and evaluating the total process of learning and teaching in terms of specific objectives, based on research in human learning . . . to bring about more

> effective instruction. (McMurrin, 1970, p. 19)

IT, then, is a systematic process that links the various components of the instructional setting; it is not an educational end in itself. Significantly, IT makes use of empirical techniques and procedures which, when properly applied in the process, generate the data educators need to develop effective instruction. Since document designers, too, are required to develop effective documents, the technology of instruction can contribute much for shaping a technology of document design.

The basic concept underlying IT is the systems analysis approach that has flourished since World War II. As weapon systems, manufacturing processes, and computer technology have become more complex and expensive, there has been a corresponding need to orchestrate systematically the different components that make up the whole "system." Systems analysis concepts and techniques enable one to break down a complex process into its components and to analyze how each contributes to the whole. Thus, the growth in popularity of such methods as input-output analysis, task analysis, decision analysis, flow charts, PERT charts, critical paths and all the other methodological accoutrements of this age of high technology.

Behavioral science's role in this movement centered on how humans and humans and machines best work together. Experimental psychologists and human factors specialists studied how to organize personnel subsystems and how to design equipment so that humans could operate them effectively and safely. Gagné (1962) and DeGreene (1970) have described a spectrum of these contributions.

The stimulant that most directly influenced IT, however, came from a different strand of behavioral science--B.F. Skinner's operant conditioning system (e.g., see Skinner, 1953). An instructional system based on Skinner's ideas makes operant reinforcement a key activity. Operant reinforcement refers to reinforcing (strengthening) specified learner behaviors so as to increase the probability of their recurrence. For learning to occur it is essential that instructors properly time and space schedules of reinforcement.

Skinner (1954, 1958) claimed that to develop a technology of instruction on the basis of operant conditioning, four things had to be known: (1) the behavior that is to be established, (2) the available reinforcers, (3) the available responses, and (4) how the reinforcements can be scheduled efficiently. The implications of this for teachers were profound. No longer was it sufficient for teachers to serve as only "presenters" of information. They now had to define and measure precisely the learning they wanted accomplished; they had to analyze their students' capabilities; and they had to draw upon the larger educational environment to support the desired behaviors where and when appropriate. In short, instruction became an instructional system where students, teachers, subject matter content, and the environment were the components to be integrated according to a prescribed plan.

Systems analysis concepts and techniques are compatible with this new view of instruction. Each component has to be broken down, described, measured, and analyzed. Programmed instruction was the first large-scale application of the systems approach to instruction; but educators soon realized that the steps involved transcended any particular mode of instruction. These steps include:

- translating instructional content into measurable objectives;
- specifying performance requirements;
- developing criterion tests that measure attainment of objectives;
- analyzing student chacteristics;
- analyzing contextual constraints and resources;

- testing and revising materials during instructional development; and
- evaluating materials against objectives.

The technology of instruction quickly developed and flourished. Glaser (1965) was among the first to articulate the systems approach, and Briggs (1966, 1968, 1970) soon after did so in detail. Mager (1962) wrote the classic book on how to translate general goals into statements of specific and measurable objectives. Glaser and Klaus (1962) were among the first to demonstrate the utility of criterion-referenced over norm-referenced testing for performance based tasks. Popham (1971) consolidated the early thinking and research on criterion-referenced measurement and outlined the basic principles for using the measures. Cunningham (1973) and Baker and Alkin (1973) outlined specific procedures and techniques for testing and revising materials during development. Scriven (1967) made useful distinctions between formative and summative evaluation in education.

IT today is firmly entrenched and accepted widely throughout much of the instructional community. The military services have institutionalized the systems approach for their training development. Many training departments in industry routinely apply IT. Schools use IT in developing system-wide curricula. There are scores of textbooks, self-instructional programs, films, and manuals on IT. Training developers and curriculum specialists share a common IT jargon and belong to the same professional associations.

A byproduct of all this interest in IT is a number of "models" that prescribe the steps for systematically developing instruction. These models vary in detail and specific steps, but they all in some way break the whole process into distinct but interrelated phases of analysis, development, and testing.

The systems orientation and prescriptive IT models are directly applicable to document design. Figure 2 depicts a process model developed by the American Institutes for Research for designing any given document. In the pre-design steps (or analysis phase) the document designer does the necessary analyses to define the purpose of the document, the audience who will use the document, and the context in which it will be used. The design steps (or development phase) begin only after all the required analyses are completed. The post-design steps (or testing phase) are done to assess whether the document achieves its aims for the intended audience.

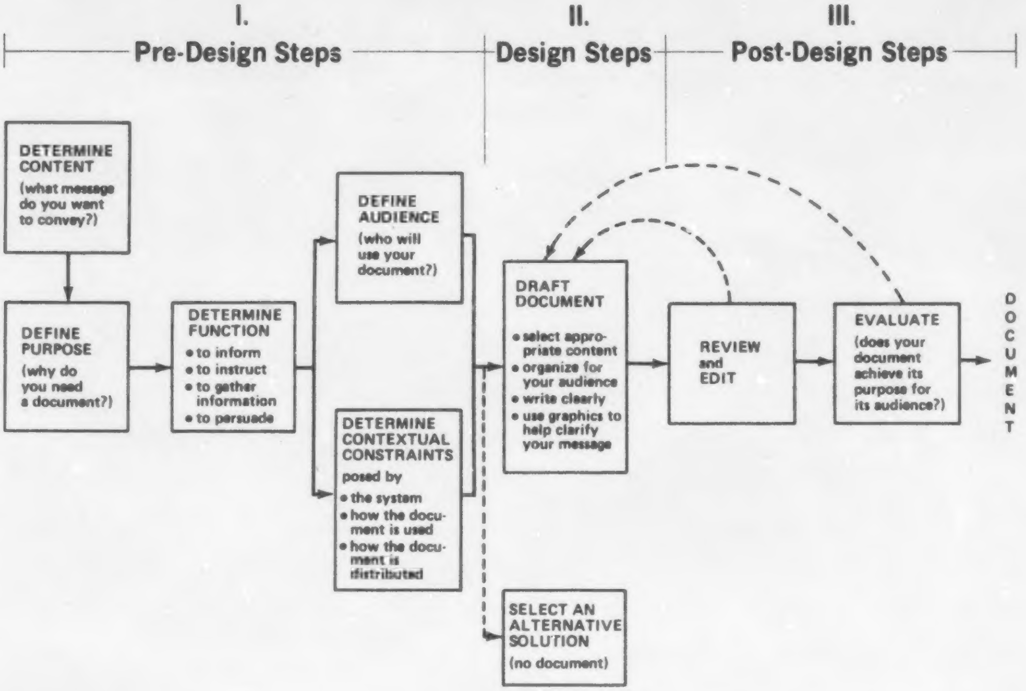
I.
Pre-Design Steps
II.
Design Steps
III.
Post-Design Steps
DETERMINE CONTENT
(what message do you want to convey?)
DEFINE PURPOSE
(why do you need a document?)
DETERMINE FUNCTION
• to inform
• to instruct
• to gather information
• to persuade
DEFINE AUDIENCE
(who will use your document?)
DETERMINE CONTEXTUAL CONSTRAINTS
posed by
• the system
• how the document is used
• how the document is distributed
DRAFT DOCUMENT
• select appropriate content
• organize for your audience
• write clearly
• use graphics to help clarify your message
SELECT AN ALTERNATIVE SOLUTION
(no document)
REVIEW and EDIT
EVALUATE
(does your document achieve its purpose for its audience?)
DOCUMENT

FIGURE 2. Document Design as a Process

Given the similarities in tasks performed by instructors, instructional developers and document designers, IT may be the one contribution from education that document designers can immediately draw upon to help them improve the quality of public documents.

# IV. Readability

## INTRODUCTION

### What is a Readability Formula?

This chapter focuses on the development and utility of readability formulas. Readability formulas are mathematical equations for predicting the level of reading ability needed to understand a particular piece of prose. They are based on *correlations* with some measure of comprehension (or with another formula); they do not specify the *causes* of problems people have in reading or understanding written material.

Readability formulas are developed by counting certain features in passages of known difficulty and correlating the frequency of each feature with the level of difficulty of the passage. A regression analysis can be used to find how much of the variance in the comprehension score is accounted for by different features.

It is theoretically possible to construct an equation for every feature and combination of features that are counted in a study. Therefore, a major research project on readability can yield many different equations. The features included in the published formulas are usually chosen as much for how easy they are to count as for their predictive value. Researchers in the field of readability strive to maximize reliability, validity, and ease of application.

### Why Develop a Readability Formula?

While psychological and psycholinguistic theories have suggested features to study in developing readability formulas, most readability research is not grounded in any theory about how people read, learn to read, or understand written information. In general, readability research is driven more by a practical need than by theory.

The need is to have a quick and easy way to assess the level of difficulty of written material. Early readability formulas were developed to measure the reading grade level of school textbooks; but in recent years the formulas have been increasingly used for many other types of material, including army training manuals, insurance policies, federal regulations, income tax instructions, and consumer credit contracts.

A score of 40 or 50 on the Flesch Reading Ease Scale (Flesch, 1948) is required in many states as a measure of the clarity of insurance policies (Pressman, 1979) and has been proposed as the standard for consumer credit contracts (MacDonald, 1979; Grogan, 1979). Specifications for military technical manuals now require use of the Recalculated Flesch Reading Ease Formula (Kincaid, Fishburne, Rogers, & Chissom, 1975) for all material written for any branch of the armed services (Kniffin, 1979).

In these cases readability formulas are being set up as the standard for measuring a writer's or editor's success in revising material that was originally too difficult for the intended audience. Readability specialists caution against using formulas as guides to rewriting (Gunning, 1952; Klare, 1979); but, given the time and budgetary constraints on most writing assignments, the temptation to write to fit the formula is great (Kern, 1979).

## Issues in Research on Readability

The new uses of readability formulas raise several important questions, including:

1. What do readability formulas really measure?
2. How well do readability formulas predict the difficulty that *adults* will have with *bureaucratic and legal documents?*
3. How well do readability formulas work as *prescriptions* for rewriting rather than as *predictors* of general level of difficulty?

4. What is the appropriate role of readability in the movement to simplify documents and what alternative standards are there for rewriting documents?

Let us review the history of readability research with these four questions in mind.

## HISTORY OF READABILITY RESEARCH

### Reviews

Chall (1958) and Klare (1963) are the best sources for an overview of the history of readability research up to the time their books were published. Both books include extensive bibliographies of research studies. Klare updated his survey of readability formulas in 1974; and recently (Klare, 1979) reviewed again, in detail, the research basis of five formulas: The Fog Index (Gunning, 1952), the Recalculated Flesch Formula (Kincaid, Fishburne, Rogers & Chissom, 1975), the FORCAST formula (Caylor, Sticht, Fox & Ford, 1973), the RIDE Scale (Carver, 1973, 1974) and the Dale-Chall (1948) readability formula.

### Early Studies

The Gray-Leary formula. Because readability formulas are mathematical equations, they can only include features of a text that can be objectively defined and counted. Many factors that influence comprehension cannot be included in a readability formula because they fail to meet these criteria. Gray and Leary (1935), authors of one of the earliest readability formulas, began by asking experts "What makes a book readable?" They gathered hundreds of different answers and divided them into four categories: content, style, format, and organization. Of all of the features in the four categories, only some of the style features could be reliably defined and measured. Gray and Leary selected 44 style variables to study and found that while 20 were significant, five of these accounted for 64 percent of the variance in people's ability to answer comprehension questions. The five variables were: the number of different hard words, the percentage of different words, the average sentence length in words, the number of prepositional phrases, and the number of personal pronouns. Increases in the first four made passages more difficult; increases in the last variable made passages easier.

McCall-Crabbs: A criterion measure. In order to develop a readability formula, one must have prose passages of known difficulty or carry out tests to find comprehension ratings for the material one wants to study. Several of the most commonly used formulas are measured against the grade level scores of the McCall-Crabbs Standard Test Lessons in Reading (McCall & Crabbs, 1925, revised in 1950, and again in 1961).

In 1925, McCall and Crabbs tested 2000 schoolchildren on 376 passages. The passages were about 150 words long and each was followed by 8 or 10 multiple-choice questions. Based on the responses, McCall and Crabbs normed each passage over many grades. That is, for a particular passage, two right answers might be equivalent to a 3.7 grade level (seventh month of the third grade); seven right answers might be equivalent to a grade level of 7.8. Most readability formulas use a 50 percent criterion. They assign the passage to the grade level at which 50 percent of the children got 50 percent of the answers correct.

The McCall-Crabbs passages were renormed in 1950 and again in 1961. Powers, Sumner and Kearl (1958) recalculated several formulas including those of Flesch, Dale-Chall, and Gunning on the basis of the 1950 McCall-Crabbs scores.

Jacobson, Kirkland and Selden (1978), however, report that in a 1976 study of 12,000 Virginia schoolchildren in grades 4, 6, 9, and 12, there was a significant difference between actual and expected scores on McCall-Crabbs passages. They conclude (p. 229) that "the 1961 norms do not accurately describe the performance of today's student population." If that study is correct, formulas based on McCall-Crabbs scores need to be recalculated again.

### Commonly Used Two-Variable Formulas

An equation with five variables like the Gray-Leary formula is difficult and time consuming to apply reliably. Fewer variables would obviously take less time and effort to count. There is a trade-off, however, between a formula's validity (correspondence to the measure of comprehension) and ease of use. The effort in readability research has been to develop formulas that maximize validity while keeping only one or two variables in the final equation.

Typically, readability formulas based on McCall-Crabbs passages have multiple correlation coefficients in the range of .60 to .70. Adding more features to the formulas does not seem to improve the predictive power significantly. However, formulas based on cloze scores (see the section on cloze tests, below) often reach correlation coefficients of .80 or .90.

Most readability formulas are based on a word factor and a sentence factor. The sentence factor is usually measured by the average number of words per sentence. The word factor is measured either by the length of the word (in letters or syllables) or by the number of words *not* on a specific list of commonly known or frequently used words.

The three most commonly used formulas include these two factors. The three formulas are the Flesch Reading Ease Scale (Flesch, 1948), the Dale-Chall Formula (Dale & Chall, 1948) and the Fog Index (Gunning, 1952, 1968).

The Flesch scale: Rudolf Flesch developed several formulas in the 1940s. The two most commonly used are the Reading Ease (R.E.) formula and the Human Interest (HI) formula (both given in Flesch, 1948). The Reading Ease formula is:

$$\text{R.E.} = 206.835 - .846\,wl - 1.015\,sl$$

where

$wl$ = number of syllables per 100 words

$sl$ = average number of words per sentence.

The equation yields a numerical score between 0 and 100. The higher the score the easier the material should be.

Flesch's Human Interest formula assumes that people will find it easier and be more motivated to read material that includes personal references and direct speech. Personal pronouns, gender-specific nouns, quotations, questions, and imperatives all raise the Human Interest score. Note that the number of personal pronouns was one of the five variables in the Gray-Leary formula and the only one of the five in which an increase improved the score.

The original R.E. formula correlated about .70 with the McCall-Crabbs criterion, which is about the same as the Dale-Chall formula (Klare, 1963). Flesch originally compared the scores to the reading level of popular magazines, not to school grade levels. However, a table is provided to convert R.E. scores to grade levels (Flesch, 1949, 1962).

The Flesch scale is one of the most commonly used formulas, possibly because it is readily available in books written for the lay person. It is also commonly used because it is so easy to apply. Flesch provides a chart that allows the user to avoid doing the mathematical calculation.

Dale-Chall formula. The Dale-Chall formula is:

$$X_{c_{50}} = .1579x_1 + .0496x_2 + 3.6365$$

where

$X_{c_{50}}$ = reading grade level of the McCall-Crabbs passage at the 50 percent criterion

$X_1$ = Dale score (the number of words in the passage not on Dale's list of 3000 words known by at least 80 percent of fourth graders)

$X_2$ = average number of words per sentence.

Based on further studies, Dale and Chall provide a correction table to get more accurate grade levels.

Klare (1979) reports that the first variable in the Dale-Chall formula (words not on the Dale list) correlated .68 with the 50 percent criterion on the McCall-Crabbs passages. With the second variable added, the multiple correlation was .70. The Dale-Chall formula is often found to correlate very highly with other formulas, even for material tested on adults (Klare, 1963; Caylor, Sticht, Fox & Ford, 1973; Carver, 1974). According to Curran (1977, p. 78) the Dale list was the highest single correlate of the comprehension measure in Bormuth's study (Bormuth, 1969; see section on Bormuth, below). Klare (1979) reports a study in which Dale-Chall came closer to matching teachers' ratings of secondary school texts than three other formulas.

Before machine counting became available, the Dale-Chall formula was difficult to apply because every word has to be checked against the Dale list. Computers, however, can handle that task easily.

The Fog Index. Robert Gunning's Fog Index is relatively easy to apply reliably by hand. The formula is

$$RGL = .4\ (ASL + \%PW)$$

where

RGL is reading grade level
ASL is average sentence length in words
and %PW is percentage of words of three or more syllables (polysyllabic words).

Klare (1979) indicates that there are no statistics on the developmental validity of Gunning's formula and Gunning's formula does not fare so well in comparative studies as the Flesch scale or Dale-Chall.

Variations of these two-variable formulas. Over a hundred readability formulas have been developed in the past fifty years. Macdonald-Ross, discussing readability as part of a more general review of language in texts, suggests that "one reason so many formulae exist is because it fits nicely into the requirements for a Ph.D. thesis!" (Macdonald-Ross, 1978, p. 236). Another reason is that formulas are developed for a specific purpose, audience, type of material, and at a particular time. As each of these factors changes, new formulas may be needed.

Powers, Sumner and Kearl (1958) used the 1950 revision of the McCall-Crabbs standards to recalculate the three formulas given in this section as well as the Farr, Jenkins and Paterson formula that is described in the following paragraph.

Other variations of these formulas have been developed for easier application. Farr, Jenkins and Paterson (1951) developed a formula in which one counts average sentence length and the number of one-syllable words. Farr, Jenkins, Paterson and England (1952) and England, Thomas and Paterson (1953) found that the new formula could be counted as reliably as and yet faster than the Flesch formula. Dunnette and Maloney (1953) found that one-syllable words can be counted in about three-fourths the time required to count all the syllables.

There has been very little other research on reliability and efficiency of using different formulas. However, new formulas and mechanisms have been developed without a research base. Fry (1968) created a graph which gives grade levels for the intersection of the values of average sentence length and the number of syllables for 100 words. The Fry graph correlates well with scores on other formulas when used on instructional materials.

McLaughlin (1969) claims that his SMOG grading can be done in less than 10 minutes, counting only polysyllabic words. Gilliland (1972) notes that SMOG usually comes out two grade levels higher than the Dale-Chall formula. Klare (1963) attributes this to the fact that McLaughlin used a 100 percent criterion on the McCall-Crabbs passages. This difference does not always hold, however. In one recent study (Morris, Thilman

& Myers, 1980), SMOG was the same or lower than Dale-Chall. It was the same on one document, one grade lower on two documents, and one and one-half grades lower on a fourth document.

Because the Dale-Chall formula continuously does very well in comparative studies, many new formulas use this as their base, with an adjusted list of acceptable words. For example, Klare (1974) reports that Brown (1965) found that Dale-Chall scores of two science books dropped two grade levels when seven science words were added to the Dale list. Similarly, Stocker (1971) adjusted the Dale list for Catholic texts by adding words known to children in Catholic schools.

The question of technical vocabulary becomes very important in assessing the readability of military, bureaucratic and legal documents. The readability score will differ, depending on whether one decides to exempt or include certain words from the list of infrequent, difficult or polysyllabic words. The Internal Revenue Service now contracts for readability scores of tax instructions, but it has determined that certain words (form, publication, schedule, etc.) are to be added to the list of easy words.

There is an untested assumption in these adjustments to the word list that, because of their experience in a subject area, readers will be familiar with a specialized vocabulary. Recent research by Hooke et al. (1979) on Air Force regulations casts doubt on that assumption. As part of a larger readability research project, Hooke gave vocabulary tests to Air Force personnel working in seven career fields. The test items were taken from regulations governing those fields. People were tested only in their own area of specialty, and yet subjects in four of the seven fields scored less than 75 percent correct on the vocabulary tests.

Writers need to be cautious about adapting a readability formula for a specialized use by deciding to exempt certain vocabulary items without conducting research to determine if, in fact, those words are well-known to the readers.

### Cloze Tests: A New Way to Measure Comprehension

Formulas developed before 1960 almost invariably used the McCall-Crabbs passages as the external source of validity. If other passages were used, they were also tested with multiple-choice comprehension tests.

Constructing a reliable and valid multiple-choice comprehension test is not a simple task. If the questions deal with detail and do not cover the main points of the passage, readers may understand the gist of the topic and still be unable to answer the questions correctly. If the questions are too difficult or are ambiguous, the reader's problem may be with the test and not with the passage. A different set of questions might yield a different score for the passage.

More recently researchers have used the cloze technique (Taylor, 1953, 1956, 1957; Klare, Sinaiko, & Stolurow, 1972) as the comprehension test in readability research. In the cloze technique, the researcher deletes every nth word (usually every fifth word) in the passage and subjects fill in the blanks. One can construct five versions of a cloze test for any passage by starting to delete at the first, second, third, fourth or fifth word. All five versions are often included in the same experiment. As Hooke et al. (1979) point out, the cloze procedure has several advantages over multiple choice tests. It is easily constructed, and, if only exact answers are accepted, a cloze test can be unambiguously scored.

There is, however, some question about what a cloze test is measuring and the best way to score a cloze test. What is the relationship between filling in blanks and the psycholinguistic processes that are involved in reading? Various scoring techniques have been tried. One suggestion is to accept synonyms as answers. Gallant (1964), in a study with primary school material, found that this did not increase the reliability or validity of the test. Another suggestion is to give the reader a choice of responses for each blank in the passage. Panackal and Heft (1978) developed a multiple choice cloze test in a two-step process. The first group was given a traditional cloze test. The most frequent responses to each blank were then given to the second group as multiple choice options. Although less reliable than the cloze forms, the multiple choice appeared to increase the test's validity. This method, however, requires two groups of subjects and two testing sessions.

In comparative studies (Bormuth, 1968, 1969; Rankin & Culhane, 1969), a score of 40-44 percent correct on a cloze test has been found to be roughly equivalent to a score of 75 percent correct on multiple-choice comprehension tests.

## The 1960s: Two Theory-Driven Studies

Following the development of the cloze technique and significant advances in psycholinguistic theory, two major new research studies in readability were undertaken during the 1960s.

Bormuth. Bormuth's work (1966, 1969), funded by the U.S. Office of Education, is the most extensive readability research project to date. Using the cloze procedure, Bormuth tested 330 passages on upper elementary and high school students. He also tested 167 variables, including all the traditional factors (syllables, letters per word), and many grammatical concepts based on linguistic theory (questions, passives, depth of embeddings, parts of speech). Bormuth considered many of his variables in relation to four different levels of discourse: the word, the minimal punctuation unit, the sentence, and the passage. Ninety-six of the 167 variables refer to passage difficulty (which is what the other readability formulas consider).

Bormuth presented many formulas, all in terms of cloze scores. His formulas are highly complex, require machine calculation, and do not yield reading grade levels directly. They are seldom used. The formulas are, therefore, of less interest than some of Bormuth's findings.

In discussing the role of linguistic variables in accounting for difficulty in whole passages, Bormuth writes, "The results would have to be described as being quite complex. That is, instead of a few underlying factors accounting for nearly all the variance, 20 factors had to be used to account for only 74 percent of the variance. And instead of all the variables exhibiting a high degree of communality, over one-fourth, 25, had common variances of .6 or less . . ." (Bormuth, 1969, p. 45).

Bormuth began the study with the hypothesis that he would find a few linguistic variables to account for the difficulty students have with texts, and that the variables would be highly interdependent. The hypothesis was not confirmed. On the contrary, as he concludes, "These studies show that any theory which sets out to explain the language comprehension processes must be far more complex than originally supposed" (Bormuth, 1969, p. 55). Bormuth's conclusion casts doubt that any of the readability formulas are adequately measuring comprehension.

Coleman. Coleman (1965, 1971) also developed new readability formulas using a cloze technique and correlating the scores with 32 variables. Most of his variables are based on how frequently exemplars of word classes are found in the text. The classification Coleman uses is derived from the work of Fries (1952). The test material was thirty-six 150-word passages covering a very broad range, from first-grade material to difficult technical prose (Miller & Coleman, 1967). In the first test with 20 college students, Coleman used the traditional cloze test with every fifth word deleted. In a second study, Miller and Coleman used two variations, deleting only one word in each passage and having subjects guess every word in succession. They then had a total of 2400 responses for each passage.

Coleman gives four readability formulas, adding another factor each time. The simplest of his formulas counts only one-syllable words:

$$\text{Cloze \%} = 1.29\ (\text{1-syl. words}) - 38.45$$

This formula predicts 74 percent of the variance. The other three formulas add, in turn, sentence length, number of pronouns and number of prepositions. The most complicated of the four formulas accounts for 83 percent of the variance. In a study to validate Coleman's formulas, Szalay (1965) found very high correlations between predicted and actual scores, ranging from .83 to .89 for the four formulas.

Liau, Bassin, Martin and Coleman (1976) recalculated the Coleman formulas using the total combined cloze scores and using computers to derive hundreds of formulas by different combinations of Coleman's 32 variables. They suggest several new formulas using number of letters and number of sentences per 100 words because these variables would be very easy for a computer to count. They found that their two-variable formula predicts 85 percent of the variance, compared to 81 percent for Coleman's original two-variable formula. They also found that the formula has a correlation of .87 with scores on Bormuth's passages. Higher predictive value is generally found in formulas that use cloze scores than in formulas based on McCall-Crabbs passages. However, Coleman's formulas, like Bormuth's, do not give grade levels directly and are not commonly used.

### The 1970s: Research in the Military

During the 1970s, much of the research in readability was conducted in or funded by the military. The Army, the Navy and the Air Force all sponsored projects in research on readability. A 1975 conference brought together researchers from the three services to review their concerns and findings (Sticht & Zapf, 1976).

The researchers were concerned about the low level of literacy in the armed services and the gap between the reading ability of the personnel and the reading level of job performance materials. The Services tried several approaches to solving this problem, including literacy training programs to raise the reading levels of poor readers, and simplifying the materials to match more closely the reading levels of military personnel. Readability formulas were considered as an aid in matching reading levels of people and manuals. However, the Services did new research on readability formulas because they had been validated with schoolchildren and college students. They were concerned that the commonly used formulas might not work for military personnel reading military documents.

Kincaid. Kincaid et al. (1975) recalculated several published formulas in a three-stage process.

(1) Subjects (Navy enlisted men) were given a standard reading test (Gates-McGintie). This allowed the researchers to assign a known reading level to each subject.

(2) The subjects were then given cloze tests on passages from relevant documents. This allowed the researchers to assign a reading grade level (RGL) to each passage. The RGL for the Kincaid formulas is the level at which 50 percent of the subjects could fill in 40 percent of the words correctly.

(3) The passages were scored for the features in each of several formulas. The recalculated Flesch Reading Ease Formula (sometimes called the Kincaid-Flesch Formula) is the one now required in the Armed Services. It is:

$$RGL = .39\ ASL + 11.80\ S - 15.59$$

where

ASL = average sentence length, in words
S = average syllables per word.

Kincaid and his colleagues also recalculated the Automated Readability Index, the Fog Index and the Fog Count

(*Guide for Air Force Writing*, 1953). Curran (1976, 1977), Klare (1979) and Kern (1979) review Kincaid's work.

FORCAST. The FORCAST formula was developed as part of an extensive study of readability in the Army. Caylor, Sticht, Fox and Ford (1973) used the same three-stage process just described, although they used a different standardized test (Nelson-Denny) and the subjects were Army recruits. They also used a 35 percent criterion instead of Kincaid's 40 percent. They found that one-syllable words correlated .86 with their cloze test criterion. In its slightly simplified version, FORCAST is very easy to calculate.

$$\text{RGL} = 20 - \frac{\text{number of 1-syl. words in a 150-word passage}}{10}$$

Caylor cautions that FORCAST was developed for a specific population and may not be valid in other circumstances.

Hooke et al. (1979) studied the reliability of the use of FORCAST for rewriting Air Force regulations and found that writers underestimated the RGL of their regulations by more than one grade level in four out of seven cases. A further examination of 37 other regulations disclosed the same trend. Hooke also notes that there was a difference between RGL of the regulation and some of the sample passages.

The RIDE Formula. Carver (1973, 1974) developed the RIDE formula (Reading-Input Difficulty Estimate) as part of a project funded by the Office of Naval Research. He used Coleman's passages and the cloze procedure but his subjects were again schoolchildren. RIDE counts only the average number of letters per word and yields a score from one to five. It does not give grade levels directly. Carver himself compared his formula to the Flesch (1948) and Dale-Chall (1948) formulas and found that they were slightly more accurate than his. The advantage of the RIDE formula is that it would be very easy to use on a computer.

## ISSUES IN THE STUDY OF READABILITY

Let us now return to the four issues raised in the first section of this chapter.

### What Do Readability Formulas Really Measure?

No one is quite sure what a readability formula measures. Curran (1976, 1977) argues that readability is not comprehensibility or useability. To the extent that readability formulas are based on atheoretical correlational studies, they tell us little about the comprehension process. Bormuth's findings also suggest that readability formulas that use only a few factors have very little validity as measures of comprehension.

Readability is necessary but not sufficient to ensure the comprehensibility of material. There are many other factors that affect whether or not written material will be understood and used than can be included in a readability formula.

Charrow (1979) and Redish (1979) explore many of the limitations of readability formulas. First of all, readability formulas assume that the text is composed of well-formed grammatical prose sentences. The attempt to use readability formulas on non-prose material (forms, lists, sections of manuals that rely heavily on illustrations or charts) is not justified by any research. A formula that counts only sentence length, syllables, or letters is not sensitive to the order of the words or complexity of the grammar. Sentences with misplaced clauses, dangling participles, or misused words will score as well on a readability formula as sentences of equal length that have none of these problems.

Gibberish, made-up words, and acronyms will cause a passage to score poorly only on a readability formula that uses a vocabulary list. Lewis Carroll's poem "Jabberwocky" receives an incredibly high 90+ on the Flesch scale. Bureaucratic and legal documents are full of made up words, acronyms and highly specialized vocabulary. Even with a readability formula that is based on a list of acceptable words, the tendency has been to add these coined words and acronyms to the list of allowed words. The research of Hooke et al. (1979) reported earlier suggests that this may not be justified.

Rothkopf (1976) points out that readability formulas do not measure the relevance of the content to the particular audience, yet instructional resesarch has shown that prior knowledge of the subject matter is an important factor in comprehension. Instructional research has also shown that students recall information better from well-organized material, and that headings and summaries aid recall. (See the chapter on instructional research for more details.) Readability formulas do not account for these important variables.

Similarly, typographic features that may affect the comprehension and use of a document are not measured by readability research. The goal of document design should be to produce documents that readers *will* use, not only ones that they *can* read when asked, assigned, or paid to in a controlled, experimental situation. Motivation is, therefore, an important issue in document design; but, again, it is not measured by readability formulas. In a recent study of ten introductory psychology textbooks, Croll and Moskaluk (1977) found that Flesch scores of the readability of the books were unrelated to students' assessments of how readable or how interesting the books are.

Psycholinguistic research has also pointed out many variables that affect comprehensibility but that are not included in readability formulas. (See the chapter on psycholinguistics in this book for more details.) Kintsch and his colleagues (Kintsch, 1974; Kintsch & Vipond, 1977; Kintsch & Van Dijk, 1978) have found that the density or number and relationships of ideas in a passage significantly affects people's ability to recall information. Subjects remembered more ideas from one passage than from another even though the passages had similar readability scores. The passages differed in the number, complexity and organization of ideas (propositions) included.

Readers use redundant information to aid comprehension (Haviland & Clark, 1974), but readability formulas favor reducing the number of words and therefore encourage the writer to eliminate redundancy. Charrow and Charrow (1979), in studying jurors' ability to understand jury instructions, found for example that sentences without relative pronouns and verbs were difficult for readers to understand (e.g., "You must not speculate to be true any *insinuation suggested* by a *question asked* a witness."). Charrow and Charrow found that their revised instructions were remembered better by jurors even when readability scores on the passages were not improved.

Although there is no research on the subject, it seems reasonable that noun strings (e.g., *human services delivery program evaluation study proposal*) are far more difficult to understand than the longer phrases that would be necessary to clarify the relationships among these nouns (*a proposal for a study to evaluate a program that delivers health services*), but the noun string will receive a better score on any of the commonly used readability formulas.

One must conclude, therefore, that readability formulas do not measure many of the features that can cause great difficulty in comprehension or use of documents.

### How Well Do Readability Formulas Predict the Difficulty that Adults Will Have with Bureaucratic and Legal Documents?

The best answer is that we do not know. There is almost no research on this topic. Most readability formulas were developed on instructional materials and tested on schoolchildren. Coleman's subjects were college students, but they are not representative of the general population; and even the highly technical passages he used are not representative of the legal and bureaucratic language found in public documents.

A considerable amount of research has been conducted in the last decade on military personnel, using military manuals, but Caylor et al.(1973) caution that it is uncertain how this research would carry over into other contexts.

The formulas have been used on bureaucratic and legal documents, but not in research studies. Bendick and Cantu (1978) compared literacy levels of welfare clients with the readability requirements of 81 documents used in various income-assistance programs. They calculated literacy levels of welfare clients by examining three data sources, (1) census data on years of schooling, (2) studies that show that high school dropouts read at lower levels than those who finish school, and (3) two small studies showing that very high percentages of welfare clients have reading skills below an eighth grade level. This evidence supports their conclusion that most welfare clients read on a level lower than eighth grade. They then used the Dale-Chall formula to calculate the reading grade requirements of the documents. Most documents scored far above the eighth grade level. Bendick and Cantu's conclusions are undoubtedly correct. Most documents written for public assistance recipients are too difficult for them to read. However, the Dale-Chall formula has never been normed on that population or type of document.

Using a readability formula to show a dramatic discrepancy between the audience's ability and the document's requirements, as Bendick and Cantu have done, may have the beneficial effect of motivating an agency to change to a simpler document. But we do not have an empirical basis for saying that if a public document scores well enough on a readability formula to match the literacy level of the clientele, it will in fact be understood by the readers.

Tax instructions from the Internal Revenue Service are now subjected to a series of computerized readability formulas with a requirement that they must score no higher than an eighth or ninth grade level on the Dale-Chall, Flesch, and Harris-Jacobson formulas. However, no research has been done to see if these formulas accurately predict the comprehension test scores of the general population of taxpayers reading income tax instructions. Carrying the formulas over from one audience and document type to another assumes a degree of similarity that may not be valid.

Even if the formulas were normed on this population and document type, there is another problem in using readability formulas to predict the difficulty that adults will have with bureaucratic and legal documents. There is a wide discrepancy in the grade levels that different formulas assign to the same documents. In a recent study of information leaflets on prescription drugs (Morris, Thilman & Myers, 1980), the researchers calculated readability levels for four versions of the same leaflet using several different formulas. They used thirteen formulas in all. Four word list formulas gave grade levels *on the average* 3.3 grades lower than formulas using syllable counts. Eliminating the word list formulas and looking only at five formulas discussed in this paper (Flesch R.E., FOG Index, SMOG, Dale-Chall, FORCAST), there is still a difference of as many as five grades between the scores using different formulas, and an average of 3.65 grades between the highest and lowest scores for each leaflet.

The formulas all predict the same rank ordering of the four leaflets. Stokes (1978) also found that the formulas he studied gave approximately the same rank ordering to a set of documents while showing differences of several grades in the actual scores.

Kern's data (Kern, 1979) also indicate that scores for the same passage using different formulas may vary by as much as four grades. Moreover, he found that the formulas do not even give the same order of difficulty to a set of passages.

Kern also looked at a related question. How accurately can a given formula predict the grade level of a passage? Chall (1958) and Powers, Sumner and Kearl (1958) estimated that the formulas they were considering, when used on school texts, would be within one grade level of accuracy. Kern did not find this to be true of the formulas used by Kincaid et al. (1975) or Caylor et al. (1973) for the passages from military documents.

Passages in the Kincaid study that were assigned to a 9-10 grade level on the basis of the cloze tests received scores ranging from grades 6 to 12 using only the Kincaid-Flesch formula. Kern concludes that "The precision of this formula is limited to predicting whether comprehensibility of the material is at or below about the tenth reading grade level." And he further argues that "A similar lack of precision is characteristic of all of the formulas /used in the Kincaid and Caylor studies/ "(Kern, 1979, p. 8).

Readability formulas can make gross distinctions between relatively simple or sophisticated text. They may be useful in making finer distinctions among instructional prose for elementary school students, because the formulas were developed on the basis of materials, readers, and reading tasks at that level. They are not fine-tuned enough to indicate reliably and validly the difference between eighth and ninth or ninth and tenth grade material, or even eighth and tenth grade material for adult readers. A computer program that gives a new reading score with every editorial change of sentence length or word substitution may be doing the document designer and the reader a great disservice.

### How Well Do Readability Formulas Work as *Prescriptions* for Rewriting Rather than as Predictors of General Level of Difficulty?

Authors of readability formulas consistently caution against using the formula as a guide to rewriting. Shortening sentences and replacing difficult words will raise the Reading Ease score of a text. It may or may not improve the comprehensibility or useability of the material.

The variables included in a readability formula are indices of difficulty, not causal factors. Klare (1979) argues that expecting comprehension to improve by writing to a readability formula is like lighting a match under a thermometer to warm up a room. The temperature on a thermometer is an index of how warm the room is. Lighting a match under the thermometer will make the index value go up, but the room will not get much warmer.

Klare (1976) reviewed 36 experimental studies that tested the effect of readability variables on readers' comprehension and/or retention. Most of the studies were doctoral dissertations, giving Klare the opportunity to look in depth at the methodologies and materials the researchers used. Only 19 of the 36 studies had statistically significant positive results (rewriting to the formula increased comprehension). However, the change in grade level in the positive studies was much greater than the change in grade level in the studies that did not show significant differences. Passages in studies with positive results were changed an average of 6.5 grades with only one case of a change less than 3 grades. Passages in studies without positive results were changed an average of 4.25 grades with two cases of changes of two grades or less. It seems that to make a difference, one has to make extreme changes in the sentence length and vocabulary.

The Navy is currently studying the effect of rewriting passages by changing only the features counted in the Kincaid-Flesch formula (Kern, 1979). In a preliminary experiment, eight passages were rewritten, once changing only vocabulary, once changing only sentence length, once changing both features. The four sets were administered under four different testing conditions to sailors at three levels of reading ability. In one of the four conditions the lowest skill group showed an increase in comprehension, but the greatest increase was not for the easiest version but for the most difficult of the rewritten passages. Except for this one instance, easier versions produced no improvement in comprehension.

Readability formulas are not meant to be prescriptions for rewriting. But the fear persists that, despite the research evidence, formulas will be used as prescriptions for rewriting. As Kern (1979, p. 18) writes: "When imposed as a standard that must be met, attention tends to be focused on meeting the score requirement rather than on organizing the material to meet the readers' informational need." This is an interesting topic on which there is, as yet, no research. Do writers who know that their work will be judged by the change in readability score make only those changes that directly affect the score? How does the availability of interactive text processing (computers that interact with the writer) affect this situation? Do writers who are working with computers that provide immediate feedback on long or difficult vocabulary words, sentence length and readability level, concentrate only on these features?

### What is the Appropriate Role of Readability in the Movement to Simplify Documents and What Alternative Standards Are There for Rewriting Documents?

The research reviewed in this paper shows that readability formulas are useful as gross measures of the difficulty of an already written text. There is no research support for the argument that reducing the grade level readability score of a document is an appropriate standard for judging how easy the document will be for people to understand and use appropriately.

Other measures, however, are more costly and time consuming to implement. As Danks (1978) and Krug (1979) point out in discussions of how to evaluate prescription drug leaflets, a true evaluation of any document would include gathering data on comprehension and behavior with an appropriate audience. The research basis and technology exist for developing comprehension tests using any of several methods: e.g., cloze, multiple-choice, free or cued recall.

The alternative to a readability formula as a prescription for rewriting is a set of guidelines and a checklist for the writer to be sure the appropriate guidelines have been applied. Post and Price (1974) developed checklists for Navy publications. Kern, Sticht, Welty and Hauke (1976) did the same for Army publications. The guidelines for Army publications were the result of an extensive research project that included an analysis of the problems in Army field manuals and technical manuals.

The advantage of guidelines and a checklist is that they can include many features of documents not counted by a readability formula. However, there are also several problems.

Do the guidelines really address the problems that cause people difficulty? With the exception of the Army research project, no extensive research has been done on the nature and extent of problems in the documents that people have to use in the real world.

What is the research basis for arguing that any specific guideline will make a difference? Post and Price (1974) give 107 guidelines grouped under ten headings for rewriting Navy technical manuals; but they do not discuss the question of research support for their guidelines. The research summarized in the other chapters of this monograph may be relevant to guidelines for document design but we have only begun the task of making the connection between research studies and the guidelines commonly found in style manuals.

It is not a trivial task to state guidelines so that independent analysts can reliably check on whether or not writers have used the guidelines appropriately. In order to measure the use of the guideline "eliminate empty words," we would need to agree on a definition or list of "empty words."

The authors of this monograph are currently involved in a research and development effort to produce reliable, validated guidelines that can be used by document designers instead of a readability formula.

# V. Human Factors

The human factors field typically has been viewed as the study of man's interaction with machines and the environment. Historically, human factors research has been concerned with the design of aircraft cockpits, consoles, controls, and displays, the study of work efficiency (e.g., time-and-motion studies in industry), and more recently with "man-in-the-system" issues. Within the last few years, human factors research has turned toward applying established methods and techniques to civil and social systems. The experimental study and design of educational systems, safety programs, industrial environments, and man-computer interactions have all become viable human factors targets.

The "methods and techniques" of human factors can be characterized as the application of experimental procedures in real world situations. Typically, although by no means universally, the concerns are for obtaining specific answers to direct questions: How big do console labels have to be in order to be read in various illuminations? What color combinations have the highest visibility? Do environmental controls such as heating and lighting affect worker performance? In a general sense, the term "human factors" could be applied to any experimental study that involves people interacting with their environment and where the results have direct, practical implications.

Public documents certainly represent one way in which people interact with their environment. Moreover, specific features of a public document -- e.g., size, legibility of print, color -- influence not only what a document user interacts with, but also how the interaction occurs and what is gained from it. The methods and techniques of human factors thus seem potentially valuable to both document design researchers and practitioners. Indeed, many of the experiments and concepts presented in other chapters of this review could be considered as falling within the domain of human factors. Some of the legibility and color research discussed in our review of typography, for example, clearly is human factors research, and some of the underlying rationales of instructional technology discussed in our review of instructional research is inspired by human factors thinking.

For this chapter we have limited our review of human factors research to those studies that were conducted for the express purpose of answering direct, applied questions and that have immediate relevance to issues in document design. Given this restriction, the literature examined did not reveal many recent studies of direct interest. In general, the studies we chose to include can be grouped into two distinct types: the "traditional," older human factors studies concerned primarily with dimensions of visual displays (e.g., labels, scales, maps, etc.), and more recent studies in which human factors implications were derived from cognitive theories and tested in applied situations.

## HEGED

The primary source of information for the "traditional" studies is Van Cott and Kinkade's (1972) Human Engineering Guide to Equipment Design (HEGED). This compendium summarizes many hundreds of studies conducted primarily between World War II and the mid-1960s. The studies vary a great deal with respect to the subjects used, types of stimuli and responses employed, and in other details. As the authors point out,

> This book was written primarily for those who must make human engineering decisions about equipment design and use. The literature on human engineering, accumulating as rapidly as it has over the past 25 years, is much too voluminous and diverse . . . to review, integrate, and translate into practical recommendations. By providing . . . authoritative data, principles, and design practices, . . . it is hoped this Guide will be a useful reference that will assist not only in identifying and resolving problems, but also in determining what is important at various design stages. It will not replace the human engineering specialist nor do his work for him. No book

> can substitute for systematic analysis or eliminate the need for research when data are needed to resolve unique and new problems. (p. viii)

The last sentence is a critical indicator of how a human factors specialist would approach document design issues, and perhaps gives a perspective on how one should view *HEGED*'s recommendations. They admit that the generality of any experimental finding will be limited by particular experimental constraints and advocate the generation of new data whenever feasible. Thus, the recommendations should be followed only when it is unfeasible to collect such data.

The recommendations presented in *HEGED* vary considerably in generality. Among the very specific recommendations which have document design implications are the following:

- For instruction decals, checklists, and labels,
    - use all capital letters;
    - letter width-to-height ratio should be 3:5;
    - height of letters should be at least .2 inches.
- In a situation in which dark adaptation will be necessary, letters should be white on a dark background.
- Colored print may be used on decals, labels, and checklists for coding purposes, but high-contrast figure-ground combinations (e.g., red-green, yellow-blue) should be chosen.
- For tables, graphs, and scales, the following graphic rules apply:
    - provide at least .166 inches between columns that are not separated by vertical rules;
    - when table columns are long, separate numbers into groups by spacing between groups of five;
    - vertical and horizontal scales on graphs should be adjusted to slightly exceed the maximum ranges of the values to be covered, thereby spreading the plotted data as widely as possible.

These recommendations in *HEGED* as well as others relevant to document design which could be mentioned, are derived from and supported by a large and diverse research literature. Some of the recommendations essentially relate to the legibility of alphanumeric information. Representative legibility research has been conducted by Atkinson, Crumley, and Lillis (1952), Brown (1953), Kappauf (1951), Long, Reid, and Queal (1951), and Tinker (1963). Other relevant document design research has been done in the general area of visual presentation considerations. Attneave (1954), Baker and Grether (1954), Chambers (1956), Foster (1964), and Siegel and Crain (1960) are a few of the many researchers reporting studies in this area. Still other research directly related to document design comes from work done in connection with the design and interpretation of data from tables, graphs, and scales. Some of the research in this area has been reported by Carter (1946), Connell (1947), Bowen and Gradijan (1963). More recently, Macdonald-Ross (1977) has summarized much of the relevant research in this area.

The Human Engineering Guide to Equipment Design also presents more general recommendations, based on the authors' extrapolation from specific findings. For example, they assert that research supports the general claim that flowcharts can produce a higher level of direct comprehension than standard prose formats for complex instructions. This conclusion was justified by the observations of specific studies that: (a) flowcharts reduce the complexity of prose needed to describe contingent relations; (b) flowcharts avoid negatives and double-negatives; (c) flowcharts distinguish relevant from irrelevant information; and (d) flowcharts reduce linguistic memory load. Importantly, these conclusions from *HEGED* generally are consistent with more recent research and practice reported on flowcharts, decision trees, and algorithms (e.g., Wright & Reid, 1973; Blaiwes, 1974; Wheatley & Unwin, 1972; Lewis & Woolfenden, 1969; Gane, Horabin, & Lewis, 1972).

## OTHER RESEARCH

Other human factors research also varies in specificity of conclusions. For example, Poulton (1967, 1969) examined variations in type size and typeface for several different tasks involving headlines and labels. He concluded that headlines are more discriminable when printed in large bold lowercase letters than when printed in uppercase letters, when the height of the letters was the same. Furthermore, if subsidiary headlines are paired with main headlines set in lowercase letters, they will have a greater attention-getting value when printed in capitals than when printed in lowercase letters of the same point size.

Poulton (1969) also found that there was a large drop in target location (specifically, for skimming food ingredient labels) when the size of the type (Univers face) was reduced from 6 point to 4 point. He recommended that the ink density for such labels and lists should have a ratio of at least .4:1; the contrast ratio between ink and paper should be at least 60%; and when two different colors of ink are used, they should have a relative ratio of 2:5:1. More generally, Poulton found that 10 point type (Times Roman) was more legible than other sizes; he recommended that a revised set of letters should be designed for every change of about 2 points in size. This latter recommendation was based on an empirical study which showed that reading rate decreased with decreasing type size.

Chapanis (1965) examined how to construct effective instructions, checklists, and labels. He concluded on the basis of his research that instructions should avoid complex or lengthy sentence structures, and that decals, checklists, and labels should be composed of common words. Chapanis makes the further important, but often overlooked, point that, for particular populations, words common to that audience should be used.

100

Again regarding very specific recommendations, several investigators have studied particular visual stimulus situations. Vartabedian (1971) examined various effects of upper- and lowercase letters on CRT displays. He found that displays of uppercase letters are searched faster than those of lowercase letters. However, he also found that text is read more rapidly and accurately in lowercase, and that with respect to isolated letters and words, uppercase is more legible. Potash (1977), in a study of the use of color on maps, found that color aids identification and search if the target color is known in advance. Also, colors as codes can be identified more accurately than size, brightness, or shape. Furthermore, redundant color improves accuracy in identification and search tasks; memory of target color is retained better than size, orientation, or shape. However, colors are affected by lighting conditions; color-blindness is a problem in about 6 percent of the population.

Potash recommended that one should choose high contrast figure-ground color combinations; choose high-differential brightness; and use as large an area of solid color as possible.

On a more general level, human factors research has provided some empirical support for existing document design principles. Thus, Kalt and Barrett (1973) found that the inclusion of advance organizers, postquestions, concrete illustrations, and delayed review (in a technical manual) led to better recall, more accurate and faster lookup scores, and more favorable attitudes in the reader. Coke and Rothkopf (1970) experimentally verified that a Flesch R.E. could predict short-term retention and reading speed. Likewise, Smith and Kinkaid (1970) demonstrated that comprehension was negatively related to words per sentence and strokes per word.

Finally, some human factors researchers concerned with graphic variations in instructional material have generated some potentially applicable results. For example, Haney (1969), in a study of instruction manual formats for troubleshooters, found that action-sequenced, tabular formats lead to reduced errors as compared to prose instructions. Similarly, Booher (1975) found that:

- pictures are better than words in communicating information about concrete objects and events;
- words are clearer in meaning than pictures and are essential in representation of abstract concepts;
- the most efficient initial understanding of procedural information occurs when pictures are used as the primary channel and printed words as a secondary, aiding channel to clarify the pictures; and
- "action" response information is better presented in print rather than pictures.

Experiments on these and similar topics blur any meaningful distinction between human factors and instructional research. Educators have long been interested in the learning effects of a variety of instructional text features, and the variables and methods of their research are similar in many respects to the interest of human factors researchers. What is studied, and how, clearly are more important considerations than is the particular discipline associated with the research.

## IN CONCLUSION

The research findings that we have presented establish quite convincingly that principles and guidelines useful to document design can be extracted from human factors research. Research on type size, or color, or tables and graphs, or letter width-to-height ratios, is applicable to the design of documents, even though the research was not done specifically on public documents. The point is that human factors research is a potential source of useful information that should not be overlooked by either document design practitioners or researchers.

But more than this, we believe that the most significant value in examining the human factors literature is not so much in the specific findings that can be excerpted, but in the perspective of the human factors scientist. The human factors "attitude" concerning careful specification of environmental influences, experimental procedures, and systematic variations of stimulus and response conditions must be an important component of the science of document design.

# VI. Typography/Graphics

Previous chapters of this review have discussed such things as how to measure how easy a public document is to read (Chapter IV), the kinds of information processing activities that occur when humans read and comprehend documents (Chapters I and II), and what aids students to learn from printed instructional materials (Chapter III). All of these issues presume that the document or instructional material is legible to begin with. However, documents can only be read or understood if they are legible. The type size must be clear and large enough to be perceived; it must be spaced adequately to be read. The document layout must highlight information, not obscure it from the reader; it must attract the reader and help him/her distinguish the different parts of the message being conveyed. This is the province of typography and graphic design.

Our use of the dual word "typography/graphics" in this chapter includes both specific and general meanings. In one sense typography has a very strict meaning: it is the *specific* way in which words, sentences, and paragraphs are actually placed on the pages of a public document. In another sense, typography in conjunction with graphic design can be viewed as encompassing the *general* design features that define the aesthetic value and overall visual effect of a document. Treating typography-plus-graphic design as one category allows us to introduce a broad field that has pervasive and obvious implications for document design.

The literature on typography/graphic design offers two basic kinds of selections: applied and theoretical. The applied literature is for the practitioner, the professional who selects a typeface or format and arranges it on the page. The practitioner is an artist and a craftsman who knows how to achieve particular aesthetic and practical effects by manipulating type, white space, or color. These practitioners can look for guidance to a variety of sources, mainly professional manuals that state the conventional wisdom accumulated over time.

The second literature is the research literature. This is the empirical work that studies the effects of various combinations of type, design, and format variables on human performance. We will examine briefly in this review the applied and theoretical literature for both typography and graphic design. The reader should note that references are made to comprehensive reviews of research done in specialized areas.

## TYPOGRAPHY

We have already noted that one kind of typographic literature is the knowledge that has accumulated from many years of practical application. This is the information essentially generated by and for practitioners. The literature consists of books and "how to" manuals that vary in the scope of typographic topics covered. Some are restricted to the rudiments of typography, some are detailed discussions of designer techniques and printing technology, and some include sophisticated issues of graphic design. Among the better known of these works are Sherbow (1916), Young (1928), Schmid (1954), Hymes (1958), Simon (1963), Biggs (1968), Craig (1971), Dair (1974), and Turnbull and Baird (1975).

The other component of the literature is typographic research. The history of typographic research goes back to the nineteenth century. Even though the nature of this research can best be described as empirical rather than scientific, a vast number of studies have been conducted. The extent of this research is indicated by the many hundreds of entries in the bibliographies of Macdonald-Ross and Smith (1973; 1977), and Tinker (1966). The research has been summarized by Tinker (1963) and Spencer (1969).

The early and most prevalent typographic research focused on factors contributing to legibility of print. These factors, to name several, include different type styles, size of type, width of line and leading, size of printed page, color of print, intercolumnar spaces, and the relationship between two or more of these factors. Tinker's (1963) research exemplifies this thrust. Some examples of this kind of research include Patterson and Tinker (1940) who examined the effects of different paint colors and background on legibility and eye movement; Tinker and Patterson (1928) and Tinker (1965) who established that people read lowercase print more easily than uppercase print. The latter finding was confirmed by Poulton (1968).

Early typographic research has had slight impact on the practices of designers, publishers, and authors. One reason for this is that the early research concerned itself with molecular issues rather than the broader context in which printed materials are used. A second and related reason is that the experimental research used simplified materials which have little correspondence with the materials actually produced by real-life document designers and printers. The products of this research thus were isolated findings with little applicability to real world concerns. Recognizing this limitation of early research, recent research (roughly since 1960) has been broadened to more complex and representative problems, including particularly those arising in the context of education and learning and those involving new technologies.

The recent typographic research typically has been concerned with the same factors as earlier research, but expanded to incorporate instructional documents and various levels of learners as well as the influences of modern printing technology. Dwyer (1971), for example, has found that color can be an effective typographic cue in a learning situation if used sparingly and if its function is explained to the learner. Coles and Foster (1975) and Rickards and August (1975) found that information set in lowercase letters is easier to read and comprehend than when printed in uppercase characters. Hartley and

Burnhill (1975) concluded that the effects of type sizes and typefaces are interactive but should be no smaller than 8 point or 10 point for instructional materials. Other researchers (Spencer et al., 1975) found that typefaces containing fine lines tend to degrade with modern printing processes. Gregory and Poulton (1970) reported that unjustified text is no more difficult to read and comprehend than justified text. Hartley, Fraser, and Burnhill (1974) provided a selective bibliography of some of the recent research done in these and other areas.

## GRAPHIC DESIGN

Graphic design is a critical part of the production of documents. The practitioner literature mainly consists of handbooks and textbooks that explain principles of design derived from experience. As a general statement, graphic designers are guided more by aesthetic and professional judgment than by scientifically verified principles. The literature reflects this artistic orientation (e.g., Bowman, 1968; Chermayeff, 1967; Chizlett, 1967).

The prevailing attitude among practitioners can be characterized as being more concerned with creativity with the intellectual or analytical processes. Design is not constrained by rigidly defined parameters of language, but only by the creative energy available to the designer for visually communicating ideas.

A number of widely accepted graphic design principles, while not derived scientifically, have grown out of the experiences of many years of practice. These focus on such aspects of typography as the overall effects of type on a page, the technical/mechanical processes of creating artwork for reproduction, and comparative examples of specific types of page layouts. Scott (1951) developed a workbook in visual design covering such topics as figure-ground relationships, color relationships, and movement. Hofmann (1965) presented design principles related to line, plane surface, space and time, paper, and design instruments (e.g., pen, crayon). Guidry and Frye (1968) developed a student oriented handbook discussing the creation of various kinds of graphs including choice of format, lettering and shading, line charts, bar graphs, and pie charts as well as tools and techniques, layout, and reproduction processes. Still other graphic designers have focused on visual effects in specific contexts, such as political posters (Barry, 1968), and marketing and advertising (Baker, 1961; Nelson, 1972).

Even though the graphic design principles and works cited above were not derived through systematic experimental investigations, research has been conducted in some areas. The bulk of this research was not originally considered as graphics research, but was done in the context of typographic, instructional, or human factors research. For purposes of the present review, we will classify all of this research as graphic design research. Lumping the research under the rubric of graphics research points out the multidisciplinary nature of document design, and demonstrates the overlap among the several research disciplines covered in this review and in this chapter.

The research has been done on what can be described loosely as "graphic devices" (Macdonald-Ross, 1978). These are alternatives to prose which typically complement the messages conveyed by prose in many public documents. Examples of graphic devices are tables, graphs and charts, as well as such supplementary format options as illustrations, color, margin width, and highlighting techniques (headings, subheadings, italics, underlining, etc.). Macdonald-Ross (1977; 1978) and Wright (1977) have reviewed the research in these areas.

The research literature on graphic devices is disparate, fragmented, extensive, and cuts across several different research disciplines. Rather than duplicate the comprehensive reviews already reported by Macdonald-Ross, we will mention several studies to illustrate the diversity and nature of the research done on graphic devices. We then will comment on its implications for document design.

A considerable amount of historical research interest centered on which type of tabular display (e.g., bar charts, line graphs) is best for presenting quantitative data. Croxton and Stein (1932) long ago tested 550 subjects on their ability to make comparative judgments in the size of bars, circles, squares, and cubes. They found that comparisons based on bar charts were more accurate than circles or squares. Culbertson and Powers (1959) examined the characteristics of bar charts and graphs that aid learning. They concluded that bar charts are easier to read, and that grouped line graphs are easier to read than segmental line graphs. Feliciano, Powers, and Kearl (1963) tested high school children and homemakers on their interpretation of data presented by text, short tables, long tables, bar charts, and assorted combinations. They found that bar charts were most effective and that performance on both bar charts and tables improved when supported by text.

Other research has been done on how best to construct specific graphic devices. One example of this is the research reported on tables, one of the most commonly used formats for presenting quantitative data. Wright (1968; 1972) and Wright and Fox (1970) distinguished between explicit tables, where information can be read directly, and implicit tables, where one must extrapolate to get the solution. Explicit tables generally were found to be superior. These investigators also concluded that tables are more efficient when users can scan them vertically.

Additional research has examined illustrations as a means of facilitating performance. Dwyer (1971; 1972) studied the effectiveness of eight types of illustrations of the heart and its parts. The illustrations ranged from simple line drawings to actual photographs of the heart, in both black and white and in color. He found that college students learned more from simple line drawings (color or black and white) than from more elaborate visuals. Borg and Schuller (1979) generally supported these results when they found that 80 soldiers learned as much about weapon systems from simple visuals as from detailed, complex visuals. In a review of the research on use of pictures with prose, Levin and Lesgold (1970) found that, for young learners at least, pictures generally facilitate learning.

Recent interest in alternative formats for presenting complex conditional instructions has spawned research on the effectiveness of algorithms (Landa, 1974; Wason, 1968). An algorithm is a form of logic tree that isolates and references decisions necessary to answer a question or solve a problem. Some of the early work tended to claim a general advantage for algorithm formats (Landa, 1965; Lewis, Horabin, & Gane, 1973; Wason, 1968; Wheatley & Unwin, 1972). Later research makes clear, however, that this advantage is task-specific (Wright & Reid, 1973; Wright, 1975, 1978).

The later research studies have probed into underlying processes, comparing algorithms in different forms and for different functions. The results are inconsistent. Wright and Reid (1973) confirmed that algorithms were better than prose in answering questions from difficult information (sentences stated indirectly or including irrelevancies), but not for straightforward information. Blaiwes (1974) found flow charts inferior even for certain kinds of difficult information. In considering task requirement, Wright and Reid (1973) found information to be memorized fared better in prose than in logical tree format. In another study, Wright (1975) found algorithms not effective for presenting instructions, when the instructions were meant to govern respondents' behavior throughout a test. Related studies (Kamman, 1975; Landa, 1974) have pointed out wide differences in difficulty between various forms of an algorithm, a factor which complicates interpreting prose-algorithm comparisons. In short, the superiority of algorithms over prose has yet to be established.

Additional research has been done on still other graphic devices. An overall criticism of the research on graphic devices is that it tends to be restricted to the study of isolated variables in rather artificial surroundings, a condition which makes it tenuous to generalize findings to real world public documents. Nevertheless, Macdonald-Ross (1977; 1978) and Hartley and Burnhill (1977) were able to offer a number of practical guidelines which have some degree of research support. While no one can claim that any of the research findings will hold for all public document situations or the entire spectrum of users, there is ample information available to improve the quality of many public documents.

# VII. A Case Study : Complex Instructions

The preceding chapters in this review of the research literature on document design are basically descriptive rather than integrative. Our approach has been to consider the research done in six different disciplines. While we have attempted to point out the relationship of each discipline to document design and where the different disciplines overlap, we have not systematically "crossed areas" in order to derive multidisciplinary document design implications. Obviously, for any particular application, such as designing a specific form or conceptualizing a particular document design research study, such a partition is highly artificial. The concern of the designer or the researcher alike is for an integrated summary of all the research relevant to the problem at hand.

We chose a descriptive rather than an integrative approach for this review for a couple of reasons. One reason stems from the "newness" of document design as an identifiable field of research. The boundaries of this emerging field are unsettled, and it is not yet clear what research literatures ought to be tapped for guidance in document design issues. The second reason is that to use an integrated approach would have required that we predetermine a set of problems (or issues) to discuss. It not only is premature to know what all the appropriate document design problems are, but this approach also runs the risk of omitting potentially critical problems. Hence, our decision to present a descriptive review.

Nonetheless, an integrated approach is feasible, at least to some degree. Any serious practitioner or researcher who is working in document design necessarily has to consider research from several disciplines concurrently. To illustrate the multidisciplinary nature of document design research, and to show what such an integrated approach to reviewing the literature could produce, we present portions of an experiment done on one particular document design problem: understanding complex conditional sentences (Holland & Rose, 1980).*

The experiment was done as part of the Document Design Project at the American Institutes for Research. The study undertook to find out what forms of complex conditionals (i.e., "If X, then Y") facilitate or hinder comprehension. The research scientists performing the study had to consider research from cognitive psychology, psycholinguistics, and information processing when they formulated their arguments and hypotheses. We reproduce below segments of their report that reflect how they drew on relevant practical and theoretical literature from these three research disciplines. The interested reader is invited to read the technical report for a complete description of the study.

◆

## UNDERSTANDING COMPLEX CONDITIONAL INSTRUCTIONS

(from Holland & Rose, 1980)

### Introduction

Documents ranging from application forms to operating manuals commonly specify actions ("submit the following") or define status (eligibility) subject to specific conditions, such as "Only disabled veterans are eligible." Frequently, these conditions are complex, consisting of many categories related by an intricate logic: "All veterans who are either employed or disabled and either single or not over 25 are eligible." Alternatively, these conditions can be specified in question form: "Are you either employed or disabled and either single or not over 25?" In order to answer the question, apply the statement, or follow the instruction, readers must first grasp the relationship--or logical structure--that links the categories, and then decide whether the condition so defined fits their particular circumstances.

---

*Holland, M. & Rose, A. *Understanding instructions with complex conditions.* Washington, D.C.: The American Institutes for Research, 1980.

Sometimes the combination of categories becomes so complex that to interpret the sentence requires a second or third reading. The complexity of this statement, taken from an individual retirement annuity brochure, is not unusual:

> You are eligible for an IRA in any given tax year if you work for yourself or someone else, full-time or part-time, and are not an active participant during any part of the tax year in any of the following types of programs: a qualified pension, profit sharing or stock bonus plan; a qualified annuity plan; a qualified bond purchase plan. . .; an annuity contract . . .; a qualified retirement plan . . .; or a government retirement plan.

Such a sentence is obviously difficult, as is confirmed by comments from users and designers of this and other forms. Whether the set of conditions is expressed by the antecedent ("if") in a conditional ("if-then") statement, or by a question or by some other construction, the problem lies not so much in the terminology as in the logical structure relating the terms. This structure must be grasped in order to interpret and apply the conditions.

Aware of such problems, specialists in written communication (e.g., Morris, 1972; Horabin & Lewis, 1974) advise rewriting or reformatting complex conditions in definitions or instructions. For example, Morris suggests: "When the [Instruction] you are working with is complex and full of conditions, use a decision table . . . [This is] a substantial improvement over . . . a narrative."

But this advice lacks both definition and support. It does not specify what type and degree of complexity call for reformatting, nor does it provide evidence that the recommended changes will work. Typically, document design guidelines such as this have neglected empirical data for defining problems and prescribing solutions.

The kind of study needed is one which will (1) diagnose the problems people have in understanding complex conditions, and (2) define those problems in terms of the extent of difficulties people have. Evidence on these questions must precede attempts to formulate and test solutions such as alternatives to prose for presenting conditional instructions.

Let us delineate the problem more precisely. We are concerned with the difficulty of multiple-category compounds which condition instructions. (These compounds may also be expressed in definitions, questions, or other kinds of sentences.) We will focus on coordinate compounds, in which the logical connectives "and" and "or" link the categories. The interesting source of difficulty at this point is the structure of the compound--the logical rule--relating the categories, rather than the content of the categories. We assume that the core task in following an instruction with compound conditions is interpreting and verifying the compound. That is, to accurately carry out the instruction "If X, do Y" requires evaluating the truth of the antecedent condition ("If X").

The following sections summarize the relevant theoretical literature, showing what is now known about the comprehension of sentences which are compounds of several categories. This summary will in turn reveal the knowledge gaps which could be empirically studied.

## The Relevant Research

One of our goals in reviewing the literature is to cover the research extending sentence comprehension from simple to coordinate structures of increasing complexity. This ground is largely unexplored, as our review will show. Because comprehension has been most thoroughly examined in the sentence verification paradigm, and because verification is the core task in the instructions we are concerned with, we begin our review with research in that paradigm. Not all the variables of interest--the structural variables inherent in coordinate compounds--have been examined in the verification paradigm. So we further cover the diverse evidence on the effects of these variables in other paradigms.

In preview, the lines of research we will survey and the questions we address are these:

*Comprehending and verifying sentences.* What is known about how comprehension and verification proceed in sentences with more than one underlying proposition? How far does this research go? What sentence variations and logical functions have been examined and what is their effect on comprehension?

*Learning logical concepts.* What is known about how logical concepts are grasped--particularly disjunction ("or") and conjunction ("and")--and their order of difficulty? Does any other work support this ordering?

*Reasoning processes.* What is known--and in relation to what sentence types--about how subjects handle the extended logical operations, such as truth table calculations, required by some sentences? Do these operations exceed what is usually considered necessary to understand a sentence? Have complex coordinate rules using multiples of "and" and "or" been examined?

*Negation in sentences.* What is the effect of negation on sentence processing generally? What is known about this effect in longer sentences with more than one negation?

*Sentence reference.* Has personal ("you are . . .") reference been investigated in sentence verification? With what result?

Comprehending and verifying sentences. Understanding a book can be characterized as a compositional process (e.g., the "semantic composition" model of Rips, Smith, and Shoben, 1978). The process involves accessing the meanings of the constituent words and combining them to yield a composite meaning for the sentence. This combining activity ordinarily consists of extracting the syntactic and logical relationships among the words of the sentence and using them as a framework for constructing sentence meaning. An intervening step in semantic composition for complex (multi-proposition) sentences is deriving the subordinate semantic units--the underlying propositions-- that compose the larger sentence.

The form which this composite meaning takes has been hypothesized in theories of memory and comprehension. Kintsch (1974) and Anderson (1976) have modeled internal representations of sentence meaning in their work on semantic memory. Clark (1973), Clark and Chase (1972), and Carpenter and Just (1975) have depicted internal sentence representations in theories of the mental processes underlying tasks such as verification. Though differing in several respects, these theories posit in common a semantic representation which encodes the sentence in abstract propositional form.* Guiding this representation is a "deep structure assumption" that each idea in the sentence is a separate underlying proposition. The structural relations (subordination, coordination) among the component propositions are preserved in the final sentence representation.

The composition process comes into play not only in representing sentence meaning but in using that representation in tasks such as question-answering and deciding truth or falsity. For example, to confirm a two-proposition sentence such as "Canaries have wings and beaks," we first evaluate the truth of the component propositions ("Canaries have wings" and

---

*A *proposition* is a unit of meaning that contains the functional relations present in a simple sentence: subject-predicate, etc.

117

"Canaries have beaks"), as well as the relationship implied by "and." From those results we derive the composite sentence truth value.

Assuming that sentence verification proceeds by just such a combination of propositions, Rips, Smith and Shoben (1978) used a verification task to investigate the order in which the components are evaluated, as well as the form in which sentence meaning is represented. They focused on sentence frames of the garden variety subject-verb-adjective-object, such as "Canaries have yellow wings." Reaction times confirmed that subjects do analyze, represent, and verify even such simple sentences propositionally.*

In probing sentences that relate three categories (canaries, wings, yellow), Rips et al. extended previous studies of the verification of simpler sentences (two-category, one-proposition sentences such as "Canaries have wings"), which have dominated research on semantic memory (e.g., Collins & Quillian, 1970; Loftus, 1973; Smith, Shoben & Rips, 1974). But all these studies limited the central variables to lexical relations (e.g., yellow legs versus yellow wings) within a single syntactic frame. What happens when we are faced with understanding and verifying longer and more diverse sentences--those which relate a greater number of categories and in which the linking relationships take the form of logical rules such as disjunction and conjunction?

There is a second line of sentence verification research that brings us closer to seeing how we process sentences with logical relationships of varying types and complexity. Begun by Wason (1961), extended and formalized by Trabasso, Rollins and Shaughnessy (1971), Clark and Chase (1972), and Carpenter and Just (1975), investigations of the mental operations in sentence comprehension have used the verification paradigm to test the effect on reaction time of variations of the sentence frame itself. These variations have included not only syntactic aspects (the active-passive and cleft transformations) and semantic aspects (such as markedness) but also, of concern to us here, the logical aspects of negation, comparison, and quantification.

---

*"Canaries have wings" and "Wings are yellow" are two propositions derived from the sentence.

All these factors have been found to affect the difficulty of verifying a sentence. To account for these effects, ad hoc comprehension models have been proposed (e.g., Clark & Chase, 1972), which formalize sentence verification as discrete operations occurring in three to four serially ordered stages. These models predict the time required for each stage and thus for the final evaluation response.

The initial stage, termed sentence encoding, transforms the stimulus sentence into a semantic representation, in abstract propositional form. A subsequent stage of verification compares this representation with information about the referent, also represented propositionally.

The comparison stage is conceived of as a compositional process, a series of matching operations which compare the components of the two representations (sentence and referent), and test for identity or discrepancy. The comparison operations contribute independently to the time needed to compute a composite truth value.

The logical operators negation and quantification complicate processing in this model by adding to the initial semantic representation of a sentence--increasing the number and depth of the propositions to be encoded.* This in turn increases the number of matching operations required to judge sentence truth or falsity.

The research stimulated by sentence verification models is limited in that it generally looks at one sentence variation at a time (e.g., affirmation-negation) restricts negation to a single occurrence, and relates no more than two concepts in a sentence. The experimental sentences have not extended beyond one or two underlying propositions in complexity. To look more closely at variations in the logical relation itself, and how these affect verification, we must consider a wider range of logical rules applied to a greater number of categories within a sentence. In particular, coordinate compounds, which use the logical connectives "and" and "or," have not been investigated in the verification paradigm.

---

*For example, "The square is not red" would be encoded as a core affirmative proposition embedded in a negative one: "Not (the square is red)."

Learning logical concepts. One branch of concept learning research* has studied the relative difficulty of logical rules relating two categories. Early research on the experimental acquisition of bi-dimensional conceptual rules, using specially constructed stimulus referents, revealed a reliable order of rule difficulty (Neisser & Weene, 1962; Haygood & Bourne, 1965; Bourne, 1967). The simple affirmative ("the item is square") was easiest, the conjunctive harder ("the item is square *and* red"), and the disjunctive ("the item is square *or* red") was the most difficult. The conjunctive-disjunctive ("and-or") difference persists in the wider research on classification processes. Subjects show consistent biases toward selecting, and greater facility in learning, conjunctive over disjunctive classifications (e.g., Bruner, Goodnow, & Austin, 1956; Hunt & Hovland, 1960). Supporting the rule-learning results, the language acquisition literature indicates that "and" precedes "or" developmentally (Braine, 1978). Further, Rose and Cox (1980) offer an intuitively appealing explanation of the general cognitive advantage of "and," by analogy to electronic circuitry.

The work on rule learning is applicable to our concern with complex conditionals. We would expect that the order of difficulty of the logical structures used in concept attainment tasks will generalize to sentence comprehension. But there are two obvious gaps in this application. First, the concept-learning paradigm is primarily non-verbal: the "and-or" distinction remains to be examined in sentence comprehension theory and research.* Second, the former paradigm has yet to extend the investigation of logical relationships from two to varying numbers of categories.

---

*A "concept" in this paradigm is defined as the relationship between categories (stimulus attributes or combinations of attributes) as characterized by a logical connective, or "rule"--e.g., "X or Y" expresses a disjunctive concept.

Reasoning processes: Processing sentences in experimental tasks such as verification can entail a complex chain of reasoning, exceeding the operations normally seen as sentence comprehension. This extended processing appears to be the case when the sentence expresses a complex relationship among categories--as reflected, for example, in determining what combination of categories or their negations makes the sentence true.

A well-studied example of extended reasoning is the conditional relationship expressed in "if-then" sentences, presented to subjects in various sorts of problem-solving tasks such as syllogisms. In order to solve the problem, subjects must go beyond understanding the given sentence--which is simple enough with the "if-then" construction in natural language--and deduce from that sentence the set of true and false states (combinations of propositions) defined by the truth table of material implication.** This deduction is notoriously difficult to make (Johnson-Laird & Tagart, 1969; Wason & Johnson-Laird, 1972; Johnson-Laird, 1975). As another example, there has been some work on quantifiers, as mentioned in our review of verification. The inordinate difficulty of using sentences with two quantifiers ("some Xs have few Ys") in a range of verification contexts (Johnson-Laird, 1970) suggests that for some functions these sentences require logical calculations beyond extracting the basic sentence meaning.

To reinforce the distinction we are drawing between comprehension and reasoning, we should note that varying the linguistic structure appears to have little effect on the logical complexity of statements of relationships such as quantification or material implication. Attempts to reformulate the language of the conditional or of doubly quanitified sentences have failed to improve substantially the reasoning errors that pervade deduction tasks (see, e.g., Legrenzi, 1970; Johnson-Laird, 1970)*.

---

*There is one exception. The Rose and Cox study (1980) suggests that for verbally presented rules, conjunction is again easier than disjunction.

**"Material implication" defines the set of true-false contingencies between two propositions connected by "if-then" (the formal logical meaning of "if-then").

We have chosen to focus on reasoning processes engaged by rules of a medium degree of complexity--as expressed by coordinate conjunctions, "or" and "and." These operations, when combined with negation in multi-category sentences, should produce the same sort of reasoning burden as conditional relations, though to a lesser degree. That is, verifying the sentence becomes an intricate process of deriving the set of possible ("true") category combinations defined by the rule, then matching those possibilities against similarly encoded reference information to determine overall truth or falsity. This process goes beyond the steps outlined in traditional sentence comprehension models, to encompass reasoning operations such as constructing a mental truth table . Such operations have been experimentally explored in the case of logically less and logically more complex structures (e.g., simple sentences on the one hand and conditional relations on the other), as we have shown. The gap is at the middle ground of conjunction and disjunction.

Braine (1978) offers a specification of what he terms "inference rule schemata," which characterize operations in natural logic--such as reasoning with "or," reasoning with "and," eliminating alternatives, detecting contradictions, and deducing the truth sets of material implication. These formulas are of help in defining the readings of "or" and "and" and in showing the conclusions to be reached from given premises. But Braine's specifications do not suggest how the schemata are combined and used, which is our chief concern here.

---

*Example: "John is taller than Bill; Jack is taller than John."

Negation in sentences. Part of our concern with complex conditions is to see how negation complicates coordinate logic. Previous research convinces us that negative sentences ("It is not A.") are harder than affirmatives ("It is A.") in most instances (see summaries of the evidence in Wason, 1962, and Trabasso, 1972). Although the difficulty disappears in certain contexts (e.g., plausible denial, Wason, 1965; change in meaning, Greene, 1970; and other situations that are natural to negation, Wales & Grieve, 1969), the general conclusion holds for the majority of experimental tasks. For example, negative sentences are harder to recall (Miller, 1962; Mehler, 1963; Clark & Card, 1969) and to match (Miller & McKean, 1964); they complicate three-term problem-solving (Clark, 1969); and they take longer to verify against pictures of semantic memory (Wason, 1959, 1961; Wason & Jones, 1963 ; Gough, 1965, 1966; Slobin, 1966; Trabasso et al., 1971; Clark & Chase, 1972; Chase & Clark, 1972; Anderson & Reder, 1974).

The negation effect is robust. In implicit form--e.g., "doubt" for "think not"--negation appears to increase difficulty fully as much as its overt expression (Sherman, 1969; Just & Clark, 1973; Just & Carpenter, 1975). This result generalizes to implicit negation in quantifiers ("few" for "not many," Just & Carpenter, 1971).

How negation varies with other factors has been studied to a limited extent: the *scope* of negation; its *position* in the sentence; and the *truth value* of the sentence. Studies of the interactions of these factors with negation are reviewed by Carpenter and Just (1975).

The number of negative words in a sentence has been looked at only indirectly, and the results are tentative. Not surprisingly, adding negations appears to create successive comprehension decrements (Sherman, 1976; Charrow & Charrow, 1978, 1979; Johnson-Laird, 1970; Legrenzi, 1970). But critical to the number of negatives in a mutiply-negated sentence is the length and complexity of the sentence as well as the structural relationship between the negative words. For example, the words may relate independently (e.g., "not X and not Y") or hierarchically (one within the scope of another, as "not both X and not Y").

The amount and the structure of negation have not been systematically manipulated. Nor has the possible interaction been tested among negation and sentence length and complexity. Finally, how these variables affect the accuracy of sentence verification and other comprehension tasks remains to be probed. Reaction time has been the primary dependent variable in traditional studies.

Sentence reference. A final relevant aspect of verifying logically complex conditions is the nature of the information against which the conditions are to be evaluated. The verification tasks we have cited, from that of Wason (1961) to that of Rips et al. (1978), typically refer the stimulus sentences to one of two kinds of items: information stored in semantic memory (e.g., Rips et al., in the tradition of Collins and Quillian) and specially constructed experimental stimuli. The latter take the form of abstract figures (e.g., Clark & Chase, 1972) or pictures of events (e.g., Gough, 1965, 1966).

Our interest is with a third type of referent--information stored in memory about oneself. Personal meaning of this sort, which varies with the identity of the reader, is one form of what linguists call deixis, or location-specific reference. It is typical of the instructions which stimulated this study ("If you are . . ., do X."). It has not been studied psycholinguistically.

Recapitulation. We have covered the theoretical background for our inquiry into instructions with logically complex conditions, as well as the relevant prior research. Quite simply, there is very little evidence for the questions we proposed earlier:

- What factors--at what levels and in what combinations--contribute to the difficulty of multicategory conditional instructions?
- How great are these difficulties, and in particular, do there exist thresholds of sentence complexity beyond which comprehension deteriorates markedly?

# References

## PSYCHOLINGUISTICS

Anderson, J. & Paulson, R. Representation and retention of verbatim information. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 439-451.

Anderson, J., & Reder, L. Negative judgments in and about semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 664-681.

Anderson, R., Reynolds, R., Schallert, D., & Goetz, E. Frameworks for comprehending discourse. *American Educational Psychology Journal*, 1977, *14*, 367-381.

Baird, R. Congruence and negative information as determinants of answers to questions of entailment. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 545-550.

Bates, E., Masling, M., & Kintsch, W. Recognition memory aspects of dialogue. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, *4*, 187-197.

Begg, I., & Clark, J. Contextual imagery in learning and memory. *Memory and Cognition*, 1975, *3*, 117-122.

Begg, I., & Paivio, A. Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 821-827.

Bever, T. A survey of some recent work in psycholinguistics. In W. Plath (Ed.), *Specification and utilization of a transformational grammar* (Sci. Rep. No. 3, Contact AF 19 (628-5127), Cambridge, Mass.: Air Force Cambridge Research Lab., 1968.

Bever, T. The cognitive basis for linguistic structures. In J. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley, 1970.

Bever, T., Garrett, M., & Hurtig, R. Interaction of perceptual processes and ambiguous sentences. *Memory and Cognition*, 1973, *1*, 277-286.

Bock, K. The effect of a pragmatic presupposition on syntactic structure in question answering. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 723-734.

Bock, J., & Brewer, W. Reconstructive recall in sentences with alternative surface structures. *Journal of Experimental Psychology*, 1974, *103*, 837-843.

Bransford, J., Barclay, J., & Franks, J. Sentence memory: A constructive versus interpretitive approach. *Cognitive Psychology*, 1972, *3*, 193-209.

Bransford, J., & Franks, J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, *2*, 331-350.

Bransford, J., & Johnson, M. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 717-726.

Brewer, W., & Lichtenstein, E. Recall of logical and pragmatic implications in sentences with dichotomous and continuous antonyms. *Memory and Cognition*, 1975, *3*, 315-318.

Broadbent, D. Word-frequency effect and response bias. *Psychological Review*, 1967, *74*, 1-15.

Brown, C. & Rubenstein, H. Test of response bias explanation of word-frequency effect. *Science*, 1961, *133*, 280-281.

Cairns, H. & Kamerman, J. Lexical information processing during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 170-179.

Carpenter, P. Extracting information from counterfactual clauses. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 512-521.

Carroll, J. Measurement properties of subjective magnitude estimates of word frequency. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 722-769.

Catlin, J., & Micham, D. Semantic representations as procedures for verification. *Journal of Psycholinguistic Research*, 1975, *4*, 209-225.

Chafe, W. *Meaning and structure of language.* Chicago: University of Chicago Press, 1970.

Champagnol, R. Genetic evolution of the reading speed of words in relation to their frequency. *Annee Psychologique*, 1971, *71*, 407-416.

Charrow, R. & Charrow, V. Making legal language understandable: Psycholinguistic study of jury instructions. *Columbia Law Review*, 1979, *79*, 1306-1374.

Charrow, V., & Charrow, R. *The comprehension of standard jury instructions: A psycholinguistic approach.* Arlington, Va.: Center for Applied Linguistics, 1978.

Charrow, V., & Charrow, R. *Characteristics of the language of jury instructions.* Paper presented at the Georgetown Roundtable on Language and Linguistics, Washington, D.C., March, 1979.

Chomsky, N. *Syntactic structures.* The Hague: Mouton, 1957.

Chomsky, N. *Aspects of the theory of syntax.* Cambridge, Mass.: M.I.T. Press, 1965

Cieutat, V., Stockwell, F., & Noble, C. The interaction of ability and amount of practice with stimulus and response meaningfulness (m,m′) in paired-associate learning. *Journal of Experimental Psychology*, 1958, *56*, 193-202.

Clark, H. Inferences in comprehension. In D. LaBerge and S. Samuels (Eds.), *Perception and comprehension*, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Clark, H., & Chase, W. On the process of comparing sentences against pictures. *Cognitive Psychology*, 1972, *3*, 472-517.

Clark, H., & Clark, E. Semantic distinctions and memory for complex sentences. *Quarterly Journal of Psychology*, 1968. *20*, 129-138.

Clark, H. & Haviland, S. Comprehension and the given-new contract. In R. Freedle (Ed.), *Discourse production and comprehension.* Hillside, N.J.: Lawrence Erlbaum Associates, 1975.

Clark, H., & Sengul, C. In search of referents for nouns and pronouns. *Memory and Cognition*, 1979, *7*, 35-41.

Clifton, C., Kurcz, I., & Jenkins, J. Grammatical relations as determinants of sentence similarity. *Journal of Verbal Learning and Verbal Behavior*, 1965, *4*, 112-117.

Clifton, C., & Odom, P. Similarity relations among certain English sentence constructions. *Psychological Monographs*, 1966, *80*, (5).

Cofer, C. (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

Cofer, C., & Musgrave, B. (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963.

Coleman, E.. The comprehensibility of several grammatical transformations. *Journal of Applied Psychology*, 1964, *43*, 186-190.

Coleman, E., & Kim, I. Comparison of several types of typography in English. *Journal of Applied Psychology*, 1961, *45*, 262-267.

Cox, W. Problem-solving as a function of abstract or concrete words. *Contemporary Educational Psychology*, 1978, *3*, 95-101.

Danks, J., & Glucksberg, S. Experimental psychology. *Annual Review of Psychology*, in press.

Danks, J., and Sorce, P. Imagery and deep structure in the prompted recall of passive sentences. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 114-117.

Deese, J. On the structure of associative meaning. *Psychological Review*, 1962, *69*, 161-175.

Deese, J. *The structure of associations in language and thought*. Baltimore: Johns Hopkins Press, 1965.

Delis, D., & Slater, A. Toward a functional theory of reduction transformation. *Cognition*, 1977, *5*, 119-132.

Dixon, T., & Horton, D. (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N.J.: Prentice Hall, 1968.

Dooling, D., & Christiaansen, R. Episodic and semantic aspects of memory for prose. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, *3*, 428-436.

Ehri, L., & Muzio, I. The influence of verb meanings on memory for adjectives. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 265-271.

Eifermann, R. Negation: A linguistic variable. *Acta Psychologica*, 1960, *17*, 258-273.

Epstein, W. Recall of word lists following learning of sentences and of anomalous and random strings. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 20-25.

Fillenbaum, S. Memory for gist: Some relevant variables. *Language and Speech*, 1966, *9*, 217-227.

Fillenbaum, S. On the use of memorial techniques to assess syntactic structures. *Psychological Bulletin*, 1970, *73*, 231-237.

Fillenbaum, S. Psycholinguistics. *Annual Review of Psychology*, 1971, *22*, 251-368.

Fischler, I., & Bloom, P. Automatic and attentional process in effects of sentence context on word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 1-20.

Fodor, J., & Garrett, M. Some syntactic determinants of sentential complexity. *Perception and Psychophysics*, 1967, *2*, 289-296.

Foss, D., Bever, T., & Silver, M. The comprehension and verification of ambiguous sentences. *Perception and Psychophysics*, 1968, *4*, 304-306.

Foss, D. & Jankins, C. Some effects of context on the comprehension of ambiguous sentences. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 577-589.

Franks, J., & Bransford, J. The acquisition of abstract ideas. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 311-315.

Frederiksen, C. Effects of context-induced processing operations on semantic information acquired from discourse. *Cognitive Psychology*, 1975, *7*, 139-166.

Freedle, R., & Carroll, J. Language comprehension and the acquisition of knowledge: Reflections. In J. Carroll and R. Freedle (Eds.) *Language comprehension of knowledge.* Washington, D.C.: Winston, 1972.

Fugue, R., & Phye, G. The effects of physical structure and semantic organization on recall of prose. *Comtemporary Educational Psychology*, 1978, *3*, 105-117.

Garrett, M., & Fodor, J. Psychological theories and linguistic constructs. In T. Dixon & D. Horton (Eds.), *Verbal Behavior and General Behavior Theory.* Englewood Cliffs, N.J.: Prentice Hall, 1968.

Garrod, S., & Sanford, A. Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior*, 1977,*16*, 177-190.

Garrod, S., & Trabasso, T. A dual-memory information processing interpretation of sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 155-167.

Geis, M. & Winograd, W. Semantic encoding and judgments of background and situational frequency for homographs. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *104*, 385-392.

Gibson, E., Bishop, C., Schiff, W., & Smith, J. Comparison of meaningfulness and pronounceability as grouping principles in the perception and retention of verbal material. *Journal of Experimental Psychology*, 1964, *67*, 173-182.

Glass, A., & Holyoak, K. The effect of some and all on reaction time for semantic decisions. *Memory and Cognition*, 1974, *2*, 436-440.

Gleitman, L., & Gleitman, H. *Phrase and paraphrase.* New York: W. W. Norton & Co., 1970.

Glucksberg, S. & Danks, J. Grammatical structure and recall: A function of the space on immediate memory or of the retention interval? *Perception and Psychophysics*, 1969, *6*, 113-117.

Gough, P. Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior*, 1965, *4*, 107-111.

Gough, P. The verification of sentences. The effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 492-496.

Green, D. The immediate processing of sentences. *Quarterly Journal of Experimental Psychology*, 1977, *29*, 135-146.

Grice, H. Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics* (Vol. 3). New York: Academic Press, 1975.

Grieve, R. Definiteness in discourse. *Language and Speech*, 1973, *16*, 365-372.

Hakes, D., & Cairns, H. Sentence comprehension and relative pronouns. *Perception and Psychophysics*, 1970, *8*, 5-8.

Hakes, D., Evans, J., & Brannon, L. Understanding sentences with relative clauses. *Memory and Cognition*, 1976, *4*, 283-290.

Halff, H., Ontony, A. & Anderson, R. A context-sensitive representation of word meanings. *Memory and Cognition* 1976, *4*, 378-383.

Hall, J. Learning as a function of word frequency. *American Journal of Psychology*, 1954, *67*, 138-140.

Harris, R. Memory and comprehension of implications and inferences of complex sentences. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 626-637.

Harris, R. *The psychology of pragmatic implication.* Unpublished manuscript, Kansas State University, 1975.

Hasher, L., Goldstein, D., & Toppino, T. Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 107-112.

Haviland, S., & Clark, J. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 512-521.

Hayes-Roth, B., & Hayes-Roth, F. The prominence of lexical information in memory representations of meaning. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 119-136.

Hayes-Roth, B., & Thorndyke, P. Integration of knowledge from text. *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 91-108.

Herriot, P. The comprehension of active and passive sentences as a function of pragmatic expectations. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 166-169.

Hogaboam, T., & Perfetti, C. Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 265-274.

Hogenraad, R. Disponsibilite et frequence du vocabulaire: Les adjectifs qualificatifs. *Annee Psychologique*, 1969, *69*, 407-419.

Holmes, V., Arwas, P., & Garrett, M. Prior context and the perception of lexically ambiguous sentences. *Memory and Cognition*, 1977, *5*, 103-110.

Hoosain, R. The processing and remembering of congruent and incongruent sentences. *Journal of Psycholinguistic Research*, 1974, *3*, 319-331.

Hornby, P. Surface structures and presupposition. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 530-538.

Howard, J. The attentional demands of negation in a memory scanning task. *Memory and Cognition*, 1975, *3*, 319-324.

Howes, D. On the relation between intelligibility and frequency of occurrence of English words. *Journal of the Acoustic Society of America*, 1957, *29*, 296-305.

Howes, D., & Solomon, R. Visual duration thresholds as a function of word probability. *Journal of Experimental Psychology*, 1952, *44*, 260-266.

Hupet, M., & LeBouedec, B. Definiteness and voice in the interpretation of active and passive sentences. *Quarterly Journal of Experimental Psychology*, 1975, *27*, 323-330.

Huttenlocher, J., & Strauss, S. Comprehension and a statement's relation to the situation it describes. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 300-304.

Huttenlocher, J., & Weiner, S. Comprehension of instructions in varying contexts. *Cognitive Psychology*, 1971, *2*, 360-385.

Jenkins, J., Mink, W., & Russell, W. Associative clustering as a function of verbal association strength. *Psychological Reports*, 1958, *4*, 127-136.

Johnson, M., Bransford, J., Nyberg, S., & Cleary, J. Comprehension factors in interpreting memory for abstract and concrete sentences. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 451-454.

Johnson, N. The psychological reality of phrase structure rules. *Journal of Verbal Learning and Verbal Behavior*, 1965. *4*, 469-475.

Johnson, N. Sequential verbal behavior. In T. Dixon & D. Horton (Eds.), *Verbal Behavior and General Behavior Theory*. Englewood Cliffs, N.J.: Prentice Hall, 1968.

Johnson-Laird, P. On understanding logically complex sentences. *Quarterly Journal of Experimental Psychology*, 1969, *21*, 1-13.

Jones, S. The effect of a negative qualifier in an instruction. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 497-501.

Just, M., & Carpenter, P. The relation between comprehending and remembering some complex sentences. *Memory and Cognition*, 1976, *4*, 318-322.

Just, M., & Clark, H. Drawing inferences from the presuppositions and implications of affirmative and negative sentences. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 21-31.

Karhunen, L. Presupposition and linguistic content. *Theoretical Linguistics*, 1974, *1-1/2*, 181-194.

Katz, A., & Denny, J. Memory load and concreteness in order of dominance effect for verbal concepts. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 13-20.

Kieras, D. *The relation of topics and themes in naturally occurring technical paragraphs* (Tech. Rep. 1). Tucson, Arizona: University of Arizona, January, 1979.

Kieras, D. *Modelling reading times in different reading tasks with a simulation model of comprehension* (Tech. Rep. 2). Tucson, Arizona: University of Arizona, March, 1979.

Kintsch, W., & Keenan, J. Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, 1973, *5*, 257-274.

Kintsch, W., & Kozminsky, E. Summarizing stories after reading and listening. *Journal of Educational Psychology*, 1977, *69*, 491-499.

Kintsch, W., & Vipond, D. *Reading comprehension and readability in educational practice and psychological theory.* Paper presented at the Conference on Memory, University of Uppsala, June, 1977.

Klare, G. The role of word frequency in readability. In J. Bormuth (Ed.), *Readability, National Conference on Research in English.* New York, 1968.

Klintman, H. Orthographic and frequency factors in verbal reaction time. *Psychological Research Bulletin*, 1968, *8*, 1-10.

Kolers, P. Specificity of operations in sentence recognition. *Cognitive Psychology*, 1975, *7*, 289-306.

Kozminsky, E. Altering comprehension: The effect of biasing titles on text comprehension. *Memory and Cognition*, 1977, *5*, 482-490.

Kuiper, N., & Paivio, S. Incidental recognition memory for concrete and abstract sentences equated for comprehensibility. *Bulletin of the Psychonomic Society*, 1977, *9*, 247-249.

Larkin, W., & Burns, D. Sentence comprehension and memory for embedded structures. *Memory and Cognition*, 1977, *5*, 17-22.

Layton, P., & Simpson, A. Surface and deep structure in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 658-664.

Lesgold, A., Roth, S., & Curtis, M. Foregrounding effects in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 291-308.

Loewenthal, K. Semantic features and communicability of words of different classes. *Psychonomic Science*, 1969, *17*, 79-80.

Loftus, E. Leading questions and the eyewitness report. *Cognitive Psychology*, 1975, *7*, 560-572.

Loftus, E., Miller, D., & Burns, H. Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, *4*, 1931.

Mackay, D. To end ambiguous sentences. *Perception and Psychophysics*, 1966, *1*, 426-436.

Mackay, D., & Bever, T. In search of ambiguity. *Perception and Psychophysics*, 1967, *2*, 193-200.

Mandler, G. Verbal learning. In G. Mandler & P. Mussen (Eds.), *New directions in psychology III.* New York: Rinehart & Winston, 1967.

Marshall, J. Unpublished doctoral dissertation, University of Oxford. Cited in I. Schlesinger, *Sentence structure and the reading process.* The Hague: Mouton, 1968.

Martin, E., & Roberts, K. Sentence length and sentence retention in the free recall learning situation. *Psychonomic Science*, 1967, *8*, 535-536.

Martin, E., Roberts, K., & Collins, A. Short-term memory for sentences. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 560-566.

Masson, M. Context and inferential cueing of sentence recall. *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 173-185.

Matthews, W. Transformational complexity and short-term recall, *Language and Speech*, 1968, *11*, 120-128.

McKoon, G. Organization of information in text memory. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 247-260.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 346-351.

Mehler, J., & Miller, G. Retroactive interference in the recall of simple sentences. *British Journal of Psychology*, 1964, *55*, 295-301.

Miller, G. Some psychological studies of grammar. *American Psychologist*, 1962, *17*, 748-762.

Miller, G. The psycholinguistics: On the new scientists of language. *Encounter*, 1964, *23*, 29-37.

Miller, G., & Isard, S. Free recall of self-embedded English sentences. *Information and Control*, 1964, *7*, 292-303.

Miller, J., & Kintsch, W. *Readability and recall of short prose passages: A theoretical analysis* (1980 forthcoming).

Miller, G., & McKean, K. A chronometric study of some relations between sentences. *Quarterly Journal of Experimental Psychology*, 1964, *16*, 297-308.

Moeser, S. Inferential reasoning in episodic memory. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 193-212.

Morton, J. A retest of the response-bias explanation of the word-frequency effect. *British Journal of Mathematical and Statistical Psychology*, 1968, *21*, 21-33.

Newstead, S. Semantic constraints and sentence perception. *British Journal of Psychology*, 1976, *67*, 165-173.

Noble, C. An analysis of meaning. *Psychological Review*, 1952, *57*, 421-430.

North, A., & Jenkins, L. Reading speech and comprehension as a function of typography. *Journal of Applied Psychology*, 1951, *35*, 225-228.

Offir, C. Recognition memory for presuppositions of relative clause sentences. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 636-643.

O'Neill, B. Definability as an index of word meaning. *Journal of Psycholinguistic Research*, 1972, *1*, 287-298.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, *76*, 241-263.

Paivio, A., & O'Neill, B. Visual recognition thresholds and dimensions of word meaning. *Perception & Psychophysics*, 1970, *8*, 273-275.

Palermo, D., & Jenkins, J. *Word association norms: Grade school through college.* Minneapolis: University of Minneapolis Press, 1964.

Pepper, S., & Prytulak, L. Sometimes frequency means seldom: Context effects in the interpretation of quantitative expressions. *Journal of Research in Personality*, 1974, *8*, 95-101.

Perfetti, C. Lexical density and phrase structure depth as variables in sentence retention. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 719-724.

Perfetti, C. Sentence retention and depth hypotheses. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 101-104.

Perfetti, C. & Goldman, S. Discourse functions of thematization and topicalization. *Journal of Psycholinguistics Research*, 1975, *4*, 257-271.

Pezdek, K., & Royer, J. The role of comprehension in learning concrete and abstract sentences. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 551-558.

Posnansky, C., & Rayner, K. Long-term recognition memory for sentences. *Acta Psycholinguistics*, 1976, *40*, 217-231.

Potter, M., Valian, V., & Faulconer, B. Representation of a sentence and its pragmatic implications: Verbal, linguistic or abstract? *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 1-12.

Potts, G. Integrating new and old information. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 305-320.

Quine, W. The problem of meaning in linguistics. In J. Fodor & J. Katz (Eds.), *The structure of language: Readings on the philosophy of language*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964.

Rader, L. *Comprehension and retention of prose: A literature review* (Tech. Rep. 108). Champaign, Illinois: University of Illinois at Urbana-Champaign, November 1978.

Richek, M. Effect of sentence complexity on the reading comprehension of syntactic structures. *Journal of Educational Psychology*, 1976, *68*, 800-806.

Rips, L. Quantification and semantic memory. *Cognitive Psychology*, 1975, *7*, 307-340.

Rohrman, N. The role of syntactic structure in the recall of English nominalizations. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 904-912.

Rosenberg, S. The recall of verbal material accompanying semantically well-integrated and semantically poorly-integrated sentences. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 732-736.

Rothkopf, E., & Coke, E. The prediction of free recall from word association measures. *Journal of Experimental Psychology*, 1961, *62*, 433-438.

Sachs, J. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, *2*, 437-442.

Savin, H., & Perchonock, E. Grammatical structure and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 1965, *4*, 348-353.

Scarborough, D., Gerard, L., & Cortese, C. Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory and Cognition*, 1979, *1*, 3-12.

Schlesinger, I. *Sentence structure and the reading process.* The Hague: Mouton, 1968.

Schwartz, D., Sparkman, J., & Deese, J. The process of understanding and judgment of comprehensibility. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 87-93.

Schwartz, S., & Witherspoon, K. Decision processing in memory: Factors influencing the storage and retrieval of linguistic and form identification. *Bulletin of Psychonomic Society*, 1974, *4*, 127-129.

Searle, J. Indirect speech acts. In P. Cole and J. Morgan (Eds.), *Syntax and semantics* (Vol. 3). New York: Academic Press, 1975.

Sherman, M. Adjectival negation and the comprehension of multiple negated sentences. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 143-157.

Slobin, D. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 219-227.

Smith, R., & Dixon, T. Effects of exposure: Does frequency determine the evaluative connotations of words? *Journal of Experimental Research in Personality*, 1971, *5*, 124-126.

Solomon, R., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology*, 1952, *43*, 195-201.

Stolz, W. A study of the ability to decode grammatically novel sentences. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 867-873.

Suci, G., & Hamacher, J. Psychological dimensions of case in sentence processing: Action roles and animateness. *Journal of Psycholinguistics*, 1972, *1*, 34-47.

Tenenbaum, A. Task-dependent effects of organization and context upon comprehension of prose. *Journal of Educational Psychology*, 1977, *69*, 528-536.

Tulving, E. Episodic and semantic memory. In T. Tulving and W. Donaldson (Eds.), *Organization and Memory.* New York: Academic Press, 1972.

Turner, E., and Rommetveit, R. Focus of attention in recall of active and passive sentences. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 543-548.

Tyler, L., & Marsen-Wilson, W. The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 683-692.

Tzeng, O. Sentence memory: Recognition and inferences. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *1*, 720-726.

Underwood, G. Contextual facilitation from attended and unattended messages. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 99-106.

Underwood, B., & Schultz, R. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1968.

Wales, R., & Marshall, J. The organization of linguistic performance. In J. Lyons & R. Wales (Eds.), *Psycholinguistic Papers*. Edinburgh: University of Edinburgh Press, 1966.

Wannemacher, J. Processing strategies in picture-sentence verification tasks. *Memory and Cognition*, 1974, *2*, 554-560.

Warren, R., Warren, N., Green, J., & Bresnick, J. Multiple semantic encoding of homophones and homographics in context biasing dominant or subordinate meanings. *Memory and Cognition*, 1978, *6*, 364-371.

Wason, P. Response to affirmative and negative binary statements. *British Journal of Psychology*, 1961, *52*, 133-142.

Wertsch, J. The influence of listener perception of the speaker on recognition memory. *Journal of Psycholinguistic Research*, 1975, *4*, 89-98.

Wilkes, A., & Alred, G. Prose and prejudice: Some effects of primary context on the immediate recall of information. *British Journal of Psychology*, 1978, *69*, 123-133.

Wright, P. Two studies of the depth hypothesis. *British Journal of Psychology*, 1960, *60*, 63-69.

Wright, P. Feeding the information eaters: Suggestions for integrating pure and applied research on language comprehension. *Instructional Science*, 1978, *7*, 249-312.

Yngve, V. A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 1960, *104*, 444-466.

Zillman, D. Rhetorical elicitation of agreement in persuasion. *Journal of Personality and Social Psychology*, 1972, *21*, 159-165.

## COGNITIVE PSYCHOLOGY

Atlas, M. *Addressing an audience: A study of expert-novice differences in writing.* (Technical Report No. 3). Pittsburgh, Pa.: Document Design Project, Carnegie-Mellon University, 1980.

Bamberger, J. *Developing a musical ear: A new experiment.* (Artificial Intelligence Memo No. 264). Cambridge, Mass.: Massachusetts Institute of Technology, 1972.

Bhaskar, R. *Problem solving in semantically rich domains.* Unpublished doctoral dissertation, Carnegie-Mellon University, 1978.

Bransford, J. *Learning, understanding, and remembering.* Belmont, Calif.: Wadsworth Publishing Company, 1979.

Bransford, J. & Johnson, M. Consideration of some problems of comprehension. IN W. Chase (Ed.), *Visual information processing*, New York: Academic Press, 1973.

Card, S. *Studies in the psychology of text editing systems.* Unpublished doctoral dissertation. Carnegie-Mellon University, 1978.

Card, S., Moran, T., & Newell, A. *The manuscript editing task: A routine cognitive skill.* Palo Alto, Calif.: Xerox Parc Report, 1977.

Chase, W. & Simon, H. Perception in chess. *Cognitive Psychology*, 1973, *4*, 55-81.

Clevinger, T. *Audience analysis.* New York: The Bobbs-Merrill Company, Inc., 1966.

Cole, R. & Jakimik, J. A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Dever, J. & Friend, F. *Performance effects of standardized information formats.* (Memo). Pisdataway, N.J.: Bell Telephone Laboratories, 1976.

Emig, J. *The composing process of twelfth graders.* Urbana, Ill.: National Council of Teachers of English, 1971.

Flower, L. & Hayes, J. Problem solving strategies and the writing process. *College English*, 1977, *39*, 449-461.

Flower, L. & Hayes, J. *A process model of composition.* (Tech. Rep. No. 1). Pittsburgh, Pa.: Document Design Project, Carnegie-Mellon University, 1980.

Gehring, R., Torlia, M. & Kimble, G. Recognition memory for words and pictures at short and long retention intervals. *Memory and Cognition*, 1976, *4*, 256-260.

Hayes, J. *Cognitive psychology: Thinking and creating.* Homewood, Ill.: The Dorsey Press, 1978.

Hayes, J. *The magic number ten: A condition of musical creativity.* (CIP Working Paper). Pittsburgh, Pa.: Carnegie-Mellon University, 1980.

Hayes, J. & Flower, L. *Writing as problem solving.* Paper presented at the annual meeting of the American Educational Research Association, 1979.

Hayes, J. & Flower, L. *Formulating sentences in writing: A protocol study.* Paper presented at the annual meeting of the American Educational Research Association, 1980a.

Hayes, J. & Flower, L. *Problem solving models of writing.* Paper presented at the annual meeting of the American Association for the Advancement of Science, 1980b.

Hayes, J. & Simon, H. Understanding complex task instruction. In D. Klahr (Ed.), *Cognition and instruction.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Hinsley, D., Hayes, J. & Simon, H. From words to equations: Meaning and representation in algebra word problems. In P. Carpenter and M. Just (Eds.), *Cognitive processes in comprehension.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Karlins, M. & Abelson, H. *Persuasion.* New York: Springer Publishing Company, Inc., 1970.

Kominsky, E. Altering comprehension: The effects of biasing titles on text comprehension. *Memory and Cognition*, 1977, *5*, 482-490.

La Berge, D. & Samuels, S. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293-323.

Lakoff, G. Structural complexity in fairy tales. *The Study of Man*, 1972, *1*, 128-190.

Lanham, R. *A handlist of rhetorical terms: A guide to students of English literature.* Berkeley, Calif.: University of California Press, 1968.

Lanham, R. *Style: An anti-textbook.* New Haven: Yale University Press, 1977.

Larkin, J., McDermott, J., Simon, D. & Simon, H. Expert and novice performance in solving physics problems. *Science*, in press.

Lindsay, P. & Norman, D. *Human information processing: An introduction to psychology.* New York: Academic Press, 1972.

Luria, A. *The mind of a mnemonist.* New York: Basic Books, 1968.

McCorduck, P. *Machines who think.* San Francisco: W.H. Freeman and Company, 1979.

Neisser, U. *Cognition and reality.* San Francisco: W.H. Freeman and Company, 1976.

O'Hare, F. *Sentence-combining: Improving student writing without formal instruction.* (Research Report No. 15). Urbana, Ill.: National Council of Teachers of English, 1973.

Reddy, D., Erman, L.D., Fennell, R.D. & Neely, R. The HEARSAY speech understanding system: An example of the recognition process. *Proceedings of the Third International Joint Conference on Artificial Intelligence*, Stanford, California, 1973.

Rumelhart, D. *Introduction to human information processing.* New York: John Wiley & Sons, 1977.

Sandberg, J. *What I learned in the second grade.* Paper presented at the Conference on English Education and Secondary School English (National Council of Teachers of English), 1979.

Scardamalia, M. Bereiter, C., & McDonald, J. *Role-taking in written communication investigated by manipulating anticipatory knowledge.* Paper presented at the biennial meeting of the Society for Research in Child Development, March, 1977.

Schank, R. & Abelson, R. *Scripts, plans, goals, and understanding.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Shaughnessy, M. *Errors and expectations.* New York: Oxford University Press, 1977.

Simon, H. *The sciences of the artificial.* Cambridge, Mass.: The M.I.T. Press, 1969.

Simon, H. Information processing modesl of cognition. *Annual Review of Psychology*, 1979, *30*, 363-396.

Wason, P. The drafting of rules. *The New Law Journal*, 1968, *6*, 548-549.

Winograd, T. A program for understanding natural language. *Cognitive Psychology*, 1972, *3*.

Winston, P. *Artificial intelligence*. Reading, Mass.: Addison-Wesley Publishing Company, 1977.

Wright, P. Writing to be understood: Why use sentences. *Applied Ergonomics*, 1971, *24*, 207-209.

Wright, P. & Reid, F. Written information: Some alternatives to prose for expressing the outcomes of complex contingencies. *Journal of Applied Psychology*, 1973, *57*, 160-166.

# INSTRUCTIONAL RESEARCH

Allen, K. Some effects of advance organizers and level of question on the learning and retention of written social studies materials. *Journal of Educational Psychology*, 1970, *61*, 333-339.

Allen, W. Intellectual abilities and instructional media design. *AV Communication Review*, 1975, *23*, 139-170.

Anderson, R. Educational psychology. *Annual Review of Psychology*, 1967, *18*, 120-164.

Anderson, R. & Biddle, W. On asking people questions about what they are reading. In G. Bower (Ed.), *The psychology of learning and motivation.* New York: Academic Press, 1975.

Anderson, R., & Faust, G. *The effects of formal strong prompts in programed instruction.* University of Illinois: Urbana, 1966.

Andre, T. Does answering higher-level questions while reading facilitate productive learning? *Review of Educational Research*, 1979, *49*, 280-318.

Ausubel, D. *The Psychology of meaningful verbal learning.* New York: Grune and Stratton, 1963.

Ausubel, D. *Educational psychology: A cognitive view.* New York: Holt, Rinehart and Winston, 1968.

Ausubel, D. In defense of advanced organizers: A reply to the critics. *Review of Educational Research*, 1978, *48*, 251-257.

Ausubel, D., & Fitzgerald, D. The role of discriminability in meaningful verbal learning and retention. *Journal of Educational Psychology*, 1961, *52*, 266-274.

Ausubel, D., & Fitzgerald, D. Organizer, general background and antecedent learning variables in sequential verbal learning. *Journal of Educational Psychology*, 1962, *53*, 243-249.

Ausubel, D., & Youssef, M. The role of discriminability in meaningful parallel learnings. *Journal of Educational Psychology*, 1963, *54*, 331-336.

Baker, E. & Alkin, M. Formative evaluation of instructional development. *AV Communication Review*, 1973, *21*, 389-418.

Barnes, B. & Clawson, E. Do advance organizers facilitate learning? Recommendations for further research based on an analysis of 32 studies. *Review of Educational Research*, 1975, *45*, 637-659.

Barron, R. *The effects of advance organizers upon the reception learning and retention of general science content (Final Report).* Department of Health, Education, and Welfare, Project No. 1B-030; Grant No. OEG-2-710030, November, 1971. (ERIC Documentation Reproduction Service No. ED 061554.)

Berliner, D., & Cohen, L. Trait-treatment interaction in learning. In F. Kerlinger (Ed.), *Review of research in education: 1.* Itasca, Illinois: F. Peacock, 1973, 58-94.

Bisbicos, E. A test of a simplified technique for implementing looping programs. *Journal of Programmed Instruction*, 1965. *3*, 15-20.

Boyd, W. Repeating questions in prose learning. *Journal of Educational Psychology*, 1975, *64*, 31-38.

Bracht, G. Experimental factors related to aptitude treatment interactions. *Review of Educational Research*, 1970, *40*, 627-645.

Brackbill, Y., Wagner, J., & Wilson, D. Feedback delay and the teaching machine. *Psychology in the Schools*, 1964, *1*, 148-156.

Bransford, J., & McCarrell, N. A sketch of a cognitive approach to comprehension: Some thoughts about understanding what it means to comprehend. In W. Weimer and D. Palermo (Eds.), *Cognition and the symbolic processes.* Hillsdale, N.J.: Laurence Erlbaum Associates, 1974.

Briggs, L. *Sequencing of instruction in relation to hierarchies of competence.* Pittsburgh: American Institutes for Research, 1968.

Briggs, L. *Handbook of procedures for the design of instruction.* Pittsburgh: American Institutes for Research, 1970.

Briggs, L., Campeau, P., Gagné, R., & May, M. *Instructional media: A procedure for the design of multi-media instruction, a critical review of research and suggestions for future research.* Pittsburgh: American Institutes for Research, 1966.

Briggs, L., & Hamilton, N. Meaningful learning and retention: Practice and feedback variables. *Review of Educational Research*, 1964, *34*, 545-558.

Bruner, J. Some theorems on instruction illustrated with reference to mathematics. *Theories of learning and instruction. The Sixty-third Yearbook of the National Society for the Study of Education*, Part I, 1964, *63*, 306-335.

Bruning, R. Effects of review and test-like events within the learning of prose materials. *Journal of Educational Psychology*, 1969, *59*, 16-19.

Bruning, R. Short-term retention of specific factual information in prose contexts of varying organization and relevance. *Journal of Educational Psychology*, 1970, *61*, 186-192.

Campeau, P. *Level of anxiety and presence or absence of feedback in programmed instruction.* Palo Alto, California: American Institutes for Research, February 1965. (Mimeo)

Carroll, J. *Learning from verbal discourse in educational media: A review of the literature.* Princeton, N. J.: Educational Testing Service, 1971. (ERIC Document Reproduction Service No. ED 058771.)

Coulson, J., & Silberman, H. Effects of three variables in a teaching machine. *Journal of Educational Psychology*, 1960, *51*, 135-143.

Craik, F., & Lockhart, R. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 671-684.

Crist, R. Overt versus covert responding and retention by sixth-grade students. *Journal of Educational Psychology*, 1966, *57*, 99-101.

Cronbach, L. The two disciplines of scientific psychology. *American Psychologist*, 1957, *12*, 671-684.

Cronbach, L., & Snow, R. *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington Publishers, Inc., 1977.

Cunningham, D. Evaluation of replicable forms of instruction. *AV Communication Review*, 1973, *21*, 351-361.

Davis, E. The effect of advance organizers on verbal learning of eighth graders. *Dissertation Abstracts International*, 1969, *29*, (7-A), 2112.

De Greene, K. *Systems psychology.* New York: McGraw-Hill, 1970.

Della-Piana, G. An experimental evaluation of programmed learning. *Journal of Educational Research*, 1962, *55*, 495-501.

Dooling, D., & Mullet, R. Locus of thematic effects in retention of prose. *Journal of Experimental Psychology*, 1973, *97*, 404-406.

Evans, J., Glaser, R., & Homme, L. An investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In A. A. Lumsdaine and R. Glaser (Eds.), *Teaching machines and programmed learning.* Washington, D.C.: National Education Association, 1960.

Evans, J., Glaser, C., & Homme, L. An investigation of "teaching machine" variables using learning programs in symbolic logic. *Journal of Educational Research*, 1962, *55*, 433-452.

Faw, H., & Waller, T. Mathemagenic behaviours and efficiency in learning from prose materials: Review, critique and recommendations. *Review of Educational Research*, 1976, *46*, 691-720.

Felker, D., & Dapra, R. Effects of question type and placement on problem solving ability from prose material. *Journal of Educational Psychology*, 1975, *67*, 380-384.

Fitzgerald, D., & Ausubel, D. Cognitive versus affective factors in the learning and retention of controversial material. *Journal of Educational Psychology*, 1963, *54*, 73-84.

Frase, L. Learning from prose material: Length of passage; knowledge of results, and position of questions. *Journal of Educational Psychology*, 1967, *58*, 266-272.

Frase, L. Effects of question location, pacing and mode upon retention of prose material. *Journal of Educational Psychology*, 1968, *59*, 244-249.

Frase, L. A structural analysis that results from thinking about text. *Journal of Educational Psychology Monograph*, 1969, *60*, 1-16.

Frase, L. Influence of sentence order and amount of higher level text processing upon reproductive and productive memory. *American Educational Research Journal*, 1970, *7*, 307-319.

Frase, L., Patrick, E., & Schumer, H. Effect of question position and frequency upon learning from text under different levels of incentive. *Journal of Educational Psychology*, 1970, *61*, 52-56.

Frase, L., & Schwartz, B. Typographical cues that facilitate comprehension. *Journal of Educational Psychology*, 1979, *71*, 197-206.

Gagné, R. (Ed.), *Psychological principles in system development*, New York: Holt, Rinehart and Winston, 1962.

Gagné, R., Rohwer, W. Instructional psychology. *Annual Review of Psychology.* Palo Alto: Annual Reviews, Inc., 1969.

Glaser, R. (Ed.) *Teaching machines and programmed learning, II.* Washington, D.C.: National Educational Association, 1965.

Glaser, R. Individuals and learning: The new aptitudes. *Educational Researcher*, 1972, *1*, 5-13.

Glaser, R. Components of a psychology of instruction: Toward a science of design. *Review of Educational Research*, 1976, *46*, 1-24.

Glaser, R., & Klaus, D. Proficiency measurement: Assessing human performance. In R. Gagné (Ed.), *Psychological principles in system development.* New York: Holt, Rinehart and Winston, 1962.

Glaser, R., & Resnick, L. Instructional psychology. *Annual Review of Psychology*, 1972, *23*, 207-276.

Glaser, R., & Taber, J. *Investigation of the characteristics of programmed learning sequences.* Pittsburgh: Programmed Learning Laboratory, University of Pittsburgh, 1961. (Offset)

Goldbeck, R., & Campbell, V. The effects of response mode and response difficulty on programmed learning. *Journal of Educational Psychology*, 1962, *53*, 110-118.

Graver, R., Means, R., & Johnsten T. The effect of subsuming concepts on student achievement on unfamiliar science learning material. *Journal of Research in Science Teaching*, 1972, *9*, 277-279.

Hartley, J., & Davis, I. Preinstructional strategies: The role of pretests, behavioral objectives, overviews, and advance organizers. *Review of Educational Research*, 1976, *46*, 239-265.

Hershberger, W., & Terry, D. Delay of self-testing in three types of programmed text. *Journal of Educational Psychology*, 1965, *56*, 22-30.

Holland, J. Research on programming variables. In R. Glaser (Ed.), *Teaching Machines and Programmed Learning, II.* Washington, D.C.: National Educational Association, 1965.

Jankins, J., & Bailey, V. Cue selection and mediated transfer in paired-associate learning. *Journal of Experimental Psychology*, 1964, *67*, 101-102.

Jerrolds, B. The effects of advance organizers in reading for the retention of specific facts. *Dissertation Abstracts*, 1968, *28*, 4532A.

Kirkwood, J. A comparative study of advance organizers in a classroom presentation in industrial arts. *Dissertation Abstracts International*, 1971, *31* (10-A), 5202-5203.

Klare, G., Mabry, J., & Gustafson, L. The relationship of patterning (underlining) to immediate retention and to acceptability of technical material. *Journal of Applied Psychology*, 1955, *39*, 40-42.

Krumboltz, J. The nature and importance of the required response in programmed instruction. Abstracted in W. Schramm (Ed.), *Research on programmed instruction: An annotated bibliography*. Washington, D.C.: U. S. Government Printing Office, 1964.

Krumboltz, J., & Bonawitz, B. The effect of receiving the confirming response in context in programmed material. *Journal of Educational Research*, 1962, *55*, (9), 472-475.

Lawton, J., & Wanska, S. Advance organizers as a teaching strategy: A reply to Barnes and Clawson. *Review of Educational Research*, 1977, *47*, 233-244.

Levine, J. Prompting and confirmation as a function of the familiarity of stimulus materials. *Journal of Verbal Learning and Verbal Behavior*, 1965, *4*, 421-424.

Mager, R. *Preparing instructional objectives.* Palo Alto: Fearon, 1962.

Mayer, R., & Bromage, B. *Different recall protocols for technical text due to advance organizers.* (Report No. 78-2). Santa Barbara, Calif.: Department of Psychology, University of California, 1978.

McConkie, G. Learning from text. In L. Shulman (Ed.), *Review of research in education* (Vol. 5). Itasca, Ill.: F. E. Peacock Publishers, Inc., 1977.

McKeachie, W. Instructional psychology. *Annual Review of Psychology*, 1974, *25*, 161-194.

McMurrin, J. *To improve learning: Report of the Commission on Instructional Technology*. Washington, D.C.: Government Printing Office, 1970.

McNeil, J. Programmed instruction as a research tool in reading: An annotated case. *Journal of Programmed Instruction*, 1962, *1*, 37-42.

Meyers, L., & Boldrick, D. Memory for meaningful connected discourse. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *1*, 584-591.

Myers, J., Pezdek, K., & Coulson, D. Effects of prose organization upon recall. *Journal of Educational Psychology*, 1973, *65*, 313-320.

Parkhurst, P. Generating meaningful hypotheses with aptitude-treatment interactions. *AV Communication Review*, 1975, *23*, 171-183.

Perlmutter, J., & Royer, J. Organization of prose materials: Stimulus, storage and retrieval. *Canadian Journal of Psychology*, 1973, *27*, 200-209.

Popham, J. (Ed.). *Criterion-referenced measurement*. Englewood Cliffs, N.J.: Educational Technology Publications, 1971.

Porter, D. *An application of reinforcement principles to classroom teaching*. Cambridge, Mass.: Laboratory for Research in Instruction, Graduate School of Education, Harvard University, 1961.

Pressey, S., & Kinzer, J. Auto-elucidation without programming. *Psychology in the School*, 1964, *1*, 359-365.

Reynolds, J. Cognitive transfer in verbal learning. *Journal of Educational Psychology*, 1966, *57*, 382-388.

Rickards, J. Adjunct postquestions in text: A critical review of methods and processes. *Review of Educational Research*, 1979, *49*, 181-196.

Rothkopf, E. The concept of mathemagenic activities. *Review of Educational Research*, 1970, *40*, 325-336.

Rothkopf, E. Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, 1966, *3*, 241-249.

Rothkopf, E. Some theoretical and experimental approaches to problems in written instruction. In J. Krumboltz (Ed.) *Learning and the educational process*. Chicago: Rand McNally, 1965.

Rothkopf, E., & Bisbicos, E. Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology*, 1967, *58*, 58-61.

Rothkopf, E., & Bloom, R. Effects of interpersonal interaction on the instructional value of adjunct questions in learning from written material. *Journal of Educational Psychology*, 1970, *61*, 417-422.

Scandura, J., & Wells, J. Advance organizers in learning abstract mathematics. *American Educational Research Journal*, 1967, *4*, 295-301.

Schramm, W. *Research on programmed instruction: An annotated bibliography*. Washington, D.C.: U. S. Government Printing Office, 1964.

Scriven, M. The methodology of evaluation. In R. Tyler, R. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. AERA monograph series on curriculum evaluation No. 1. Chicago: Rand McNally & Co., 1967.

Shavelson, R., Berliner, D., Ravitch, M., & Loeding, D. Effects of position and types of question on learning from prose material: Interaction of treatments with individual differences. *Journal of Educational Psychology*, 1974, *66*, 40-48.

Skanes, G., Sullivan, A., Rowe, E., & Shannon, E. Intelligence and transfer: Aptitude by treatment interactions. *Journal of Educational Psychology*, 1974, *66*, 563-568.

Skinner, B. *Science and human behavior*. New York: Macmillan Company, 1953.

Skinner, B. The science of learning and the art of teaching. *Harvard Educational Review*, 1954, *24*, 86-97.

Skinner, B. Teaching machines. *Science*, 1958, *129*, 969-977.

Stolurow, L., & Lippert, H. Prompting, confirmation and overlearning in the automated teaching of a sight vocabulary. In J. DiCicco (Ed.), *Educational Technology*. New York: Holt, Rinehart and Winston, 1964.

Tallmadge, G., & Shearer, J. Relationships among learning styles, instructional methods, and the nature of learning experiences. *Journal of Educational Psychology*, 1969, *60*, 222-230.

Tallmadge, G., & Shearer, J. Interactive relationships among learner characteristics, types of learning, instructional methods, and subject matter variables. *Journal of Educational Psychology*, 1971, *62*, 31-38.

Tobias, S. Achievement treatment interactions. *Review of Educational Research*, 1976, *46*, 61-74.

Watts, G., & Anderson, R. Effects of three types of inserted questions on learning from prose. *Journal of Educational Psychology*, 1971, *62*, 387-394.

Weiss, W., & Fine, B. Stimulus familiarization as a factor in ideational learning. *Journal of Educational Psychology*, 1956, *47*, 118-124.

Williams, J. Comparison of several response modes in a review program. *Journal of Educational Psychology*, 1963, *54*, 253-260.

Wittrock, M. Effect of certain sets upon complex verbal learning. *Journal of Educational Psychology*, 1963, *54*, 85-88.

Wittrock, M., & Lumsdaine, A. Instructional psychology. *Annual Review of Psychology*, 1977, *28*, 417-459.

Wolk, S. The influence of meaningfulness upon intentional and incidental learning of verbal material. *Memory and Cognition*, 1974, *2*, 189-193.

## READABILITY

Bendick, M., Jr., & Cantu, M.G. *The literacy of welfare clients.* Washington, D.C.: The Urban Institute, 1978. (Reprinted from *Social Science Review*, 1978, *52*.)

Bormuth, J.R. Readability: A new approach. *Reading Research Quarterly*, 1966, *1*, 79-132.

Bormuth, J.R. Cloze test readability: Criterion scores. *Journal of Educational Measurement*, 1968, *5*, 189-196.

Bormuth, J.R. *Development of readability analyses.* (Univ. of Chicago Final Rep., Project No. 7-0052, Contract No. OEC-3-7-070052-0326). Washington, D.C.: U.S. Office of Education, 1969.

Brown, W.R. Science textbook selection and the Dale-Chall formula. *School Science and Mathematics*, 1965, *65*, 164-167.

Carver, R.P. *Measuring prose difficulty: The RIDE scale.* Unpublished manuscript, 1973.

Carver, R.P. *Improving reading comprehension: Measuring readability.* (AIR Final Report, Contract No. N00014-72-C0240, Office of Naval Research.) Silver Spring, Maryland: American Institutes for Research, May 14, 1974.

Caylor, J.S., Sticht, T.G., Fox, L.C., and Ford, J.P. *Methodologies for determining reading requirements of military occupational specialties.* (HumRRO Tech. Rep. 73-5). Presidio of Monterey, California: Human Resources Research Organization, March, 1973.

Chall, J.S. *Readability: An appraisal of research and application.* Columbus, Ohio: Ohio State University, 1958.

Charrow, R. & Charrow, V. Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review*, 1979, *79*, 1308-1374.

Charrow, V. *Let the rewriter beware.* Unpublished manuscript, 1979. (Available from Document Design Center, American Institutes for Research, 1055 Thomas Jefferson, Street, N.W., Washington, D.C. 20007).

Clark, H.H. & Haviland, S.E. Psychological processes as linguistic explanation. In D. Cohen (Ed.), *Explaining linguistic phenomena.* Washington, D.C.: Winston, 1974.

Coleman, E.B. Developing a technology of written instruction: Some determiners of the complexity of prose. In E.Z. Rothkopf & P.E. Johnson (Eds.), *Verbal learning research and the technology of written instruction.* New York: Teachers College Press, Columbia University, 1971.

Croll, W. & Moskaluk, S. Should Flesch counts count? *Teaching of Psychology*, 1977, *4*, 48-49.

Curran, T.E. Readability research in the Navy. In T.G. Sticht & D.W. Zapf (Eds.), *Reading and readability research in the armed services.* (HumRRO Tech. Rep. 76-4). Alexandria, Va.: Human Resources Research Organization, 1976.

Curran, T.E. *Survey of technical manual readability and comprehensibility.* (Technical Report 77-37). San Diego, California: Navy Personnel Research and Development Center, June, 1977.

Dale, E. & Chall, J.S. A formula for predicting readability. *Educational Research Bulletin*, January 21 and February 17, 1948, *27*, 11-20 and 37-54.

Danks, J. *Communication and evaluation of patient package inserts.* Paper presented at the Conference on Patient Package Inserts, Food and Drug Administration, Washington, D.C. December, 1978.

Dunnette, M.D. & Maloney, P.W. Factorial analysis of the original and the simplified Flesch Reading Ease formulas. *Journal of Applied Psychology*, 1953, *37*, 107-110.

England, G.W., Thomas, M., and Paterson, D.G. Reliability of the original and simplified Flesch Reading Ease formulas. *Journal of Applied Psychology*, 1953, *37*, 111-113.

Farr, J.N., Jenkins, J.J., Paterson, D.G., and England, G.W. Reply to Klare and Flesch re "Simplification of Flesch Reading Ease Formula." *Journal of Applied Psychology*, 1952, *36*, 55-57.

Flesch, R.F. A new readability yardstick. *Journal of Applied Psychology*, 1948, *32*, 221-233.

Flesch, R.F. *The art of readable writing.* New York: Harper & Row, 1949.

Flesch, R.F. *The art of readable writing.* New York: Collier Books, 1962.

Fries, C.C. *The structure of English.* New York: Harcourt, Brace, 1952.

Fry, E.A. A readability formula that saves time. *Journal of Reading*, 1968, *2*(7), 513-516.

Gallant, R.M. *An investigation of the cloze tests as a measure of readability of materials for primary grades.* Unpublished doctoral dissertation, Indiana University, 1964.

Gilliland, J. *Readability.* London: Hodder & Stoughton, 1972.

Gray, W.S. & Leary, B.E. *What makes a book readable.* Chicago: University of Chicago Press, 1935.

Grogan, L. *The status of plain language legislation in the fifty states and President Carter's Executive Order 12044* Unpublished master's thesis, Rutgers, The State University of New Jersey, 1979.

*Guide for Air Force Writing.* (Air Force Manual 11-3). Maxwell, Ala.: Department of the Air Force, Maswell Air Force Base, Air University, June, 1953.

Gunning, R. *The technique of clear writing.* New York: McGraw-Hill, 1952. Revised Edition, 1968.

Harris, A.J. & Jacobson, M.D. *Revised Harris-Jacobson readability formulas.* Bethesda, Md.: College Reading Assoc., October 1974.

Haviland, S.E. & Clark, H.H. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*.

Hooke, L.R., DeLeo, P.J. & Slaughter, S.L. *Readability of Air Force publications: A criterion referenced evaluation.* (AFHRL-TR-79-21). Lowry AFB, Colo.: Technical Training Division, Air Force Human Resources Laboratory, May 1979.

Jacobson, M.D., Kirkland, C.E. & Selden, R.W. An examination of the McCall-Crabbs standard test lessons in reading. *Journal of Reading*, 1978, 224-230.

Kern, R.P. *Usefulness of readability formulas for achieving Army readability objectives: Research and state-of-the-art applied to the Army's problem.* Fort Benjamin Harrison, Ind.: Technical Advisory Service, U.S. Army Research Institute, 1979.

Kern, R., Sticht, T., Welty, D. & Hauke, R. *Guidebook for the development of Army training literature.* (HUMRRO Tech. Rep. 75-3). Alexandria, Va.: Human Resources Research Organization, 1976.

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., and Chissom, B.S. *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel.* (Research Branch Report 8-75). Millington, Tenn.: Naval Air Station Memphis, February, 1975.

Kintsch, W. *The representation of meaning in memory.* New York: Wiley, 1974.

Kintsch, W. & van Dijk, T.A. Toward a model of text comprehension and production. *Psychological Review*, 1978, *85*, 363-394.

Kintsch, W. & Vipond, D. *Reading comprehension and readability in educational practice and psychological theory.* Paper presented at the Conference on Memory, University of Uppsala, June 1977.

Klare, G.R. *The measurement of readability.* Ames, Iowa: Iowa State University Press, 1963.

Klare, G.R. Assessing readability. *Reading Research Quarterly*, 1974, *10*(1), 62-102.

Klare, G.R. A second look at the validity of readability formulas. *Journal of Reading Behavior*, 1976, *8*, 129-152.

Klare, G.R. *Readability standards for army-wide publications.* (Evaluation Report 79-1). Fort Benjamin Harrison, Ind.: U.S. Army Administrative Center, 1979.

Klare, G.R., Simaiko, H.W. & Stolurow, L.M. The cloze procedure: A convenient readability test for training materials and translations. *International Review of Applied Psychology*, 1972, *21*, 77-106.

Kniffin, J.D. The new readability requirements for military technical manuals. *Technical Communication*, Third Quarter, 1979.

Krug, R. *Evaluating documents: The case of patient package inserts.* (Tech. Rept. 2). Washington, D.C.: American Institutes for Research, 1979.

Liau, T.L., Bassin, C.B., Martin, C.J. & Coleman, E.B. Modification of the Coleman readability formulas. *Journal of Reading Behavior*, 1976, *8*, 381-386.

MacDonald, D.A. *Drafting documents in plain language.* New York: Practising Law Institute, 1979.

Macdonald-Ross, M. Language in texts: The design of curricular materials. In L.S. Shulman (Ed.), *Review of Research in Education* (Vol. 6). Itasca, Ill.: F.E. Peacock Publishers, Inc., 1978.

McCall, W.A., & Crabbs, L.M. *Standard test lessons in reading.* New York: Bureau of Publications, Teachers College, Columbia University, 1925; Revised Edition, 1950; Second Revised Edition, 1961.

McLaughlin, G.H. SMOG grading--a new readability formula. *Journal of Reading*, 1969, *12*, 639-646.

Miller, G.R. & Coleman, E.B. A set of 36 passages calibrated for comprehensibility. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 851-854.

Morris, L.A., Thilman, D. & Myers, A.M. *Application of the readability concepts to patient-oriented drug information.* 1980 forthcoming. (Now available from first author, HFD-107, 5600 Fishers Lane, Rockville, MD 20857.)

Panackal, A.A. & Heft, C.S. Cloze technique and multiple choice technique: Reliability and validity. *Educational and Psychological Measurement*, 1978, *38*, 917-932.

Post, T., & Price, H. *Requirements and criteria for improving reading comprehension of technical manuals.* (N00024-74-C-5426). Falls Church, Va.: BioTechnology, Inc., 1974.

Powers, R.D., Sumner, W.A., and Kearl, B.E. A recalculation of four adult readability formulas. *Journalism Quarterly*, 1958, *49*, 99-105.

Pressman, R. *Legislative and regulatory progress on the readability of insurance policies.* Unpublished manuscript, 1979. (Available from the Document Design Center, American Institutes for Research, 1055 Thomas Jefferson Street, N.W., Washington, D.C. 20007.)

Rankin, E.F. & Culhane, J.W. Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading*, 1969, *13*, 193-198.

Redish, J.C. *Readability*. Unpublished manuscript, 1979. (Available from the Document Design Center, American Institutes for Research, 1055 Thomas Jefferson Street, N.W., Washington, D.C. 20007.)

Rothkopf, E.Z. Comments on the paper by Siegel. In T.G. Sticht & D.W. Zapf (Eds.), *Reading and readability research in the armed services.* (HumRRO Tech. Rep. 76-4). Alexandria, Va.: Human Resources Research Organization, 1976.

Sticht, T.G. & Zapf, D.W. (Eds.), *Reading and readability research in the armed services.* (HumRRO Tech. Rep. 76-4). Alexandria, Va.: Human Resources Research Organization, 1976.

Stocker, L.P. Increasing the precision of the Dale-Chall formula. *Reading Improvement*, 1971, *8*, 87-89.

Szalay, T.G. Validation of the Coleman readability formulas. *Psychological Reports*, 1965, *17*, 965-966.

Taylor, W.L. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 1953, *30*, 415-433.

Taylor, W.L. Recent developments in the use of cloze procedure. *Journalism Quarterly*, 1956, *33*, 42-48.

Taylor, W.L. "Cloze" readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology*, 1957, *41*, 19-26.

## HUMAN FACTORS

Atkinson, W., Crumley, L., & Willis, M. *A study of the requirements of letters, numbers, and markings to be used on transilluminated aircraft control panels: V. The comparative legibility of three fonts for numerals* (Report No. TED-Namel-609-5). Philadelphia, Penn.: Naval Air Materiel Center, 1952.

Attneave, F. Some informational aspects of visual perception. *Psychological Review*, 1954, *61*, 183.

Baker, C., & Grether, W. *Visual presentation of information* (WADC TR 54-160), WPAFB, Ohio: Wright Air Development Center, 1954.

Blaiwes, A. Formats for presenting procedural instructions. *Journal of Applied Psychology*, 1974, *59*, 683-686.

Booher, H. Relative comprehensibility of pictorial information and printed words in proceduralized instructions. *Human Factors*, 1975, *17*, 266-277.

Bowen, H., & Gradijan. *Graphical display of multiparametric information. Part II, Experimental studies of chart design* (AMRL TDR-62-115). WPAFB, Ohio: Aerospace Medical Research Laboratory, 1963.

Brown, F. *A study of the requirements for letters, numbers, and markings to be used on transilluminated aircraft control panels: IV. Legibility of uniform stroke capital letters as determined by size and height to width ratio and as compared to Garamond Bold* (Report No. TED-NAMEL-609-4). Philadelphia, Penn.: Naval Air Materiel Center, 1953.

Carter, L. *A study of the best design of tables and graphs used for presenting numerical data* (Report TSEAA-694-1C). WPAFB, Ohio: Wright Air Development Center, 1946.

Chambers, E. *Design and use of scales, dials and indicators* APU Report No. 254/56). Cambridge, England: Medical Research Council, 1956.

Chapanis, A. Words, words, words. *Human Factors*, 1965, *7*, 1-17.

Coke, E., & Rothkopf, E. Note on a simple algorithm for a computer-produced reading ease score. *Journal of Applied Psychology*, 1970, *54*, 208-210.

Connell, S. *The relative effectiveness of presenting numerical data by the use of scales and graphs* (WADC, TSEAA-694-1M). WPAFB, Ohio: Wright Air Development Center, 1947.

Foster, H. *Information displays and information processing tasks* (SP-1811). System Development Corp., 1964.

Gane, C., Horabin, I., & Lewis, B. The simplification and avoidance of instruction. In I. Davies & J. Hartley, *Contributions to an Educational Technology.* London: Butterworths, 1972.

Haney, R. Functional testing performance. *Human Factors*, 1969, *11*, 181-188.

Kalt, N., & Barrett, K. Facilitation of learning from a technical manual: An exploratory investigation. *Journal of Applied Psychology*, 1973, *58*, 357.

Kappauf, W. *Design of instrument scales for maximum legibility: V. Origin location, scale break, number location, and contrast direction.* USAF Aero Med. Lab. AF-TR-6366, 1951, WPAFB, Ohio

Lewis, B., & Woolfenden, P. *Algorithms and logical trees: A self-instructional course.* Cambridge, England: Algorithms Press, 1969.

Long, E., Reid, L., & Queal, R. Factors determining the legibility of letters and words derived from elemental printers (Wade Tech. Reports Nos. 5922-5924), 1951.

Macdonald-Ross, M. Graphics in text. In L. Schulman (Ed.), *Review of Research in Education*(Vol. 5), Itasca, Ill.: F. E. Peacock Publishers, Inc., 49-85.

Potash, L. Design of maps and map-related research. *Human Factors*, 1977, *19*, 139-150.

Poulton, E. Skimming lists of food ingredients printed in different sizes. *Journal of Applied Psychology*, 1963, *53*, 52-229.

Poulton, E. Searching for newspaper headlines printed in capitals or lower-case letters. *Journal of Applied Psychology*, 1967, *51*, 417.

Siegel, A., & Crain, K. Experimental investigations of cautionary signal presentations. *Ergonomics*, 1960, *3*, 339-356.

Smith, E., & Kincaid, J. Derivation and validation of the automated readability index for use with technical materials. *Human Factors*, 1970, *12*, 457-464.

Tinker, M. *Legibility of print.* Ames, Iowa: Iowa State University Press, 1963.

Van Cott, H., & Kinkade, R. *Human engineering guide to equipment design.* New York: McGraw-Hill, 1972.

Vartabedian, A. The effect of letter size, case, and generation method on CRT display search time. *Human Factors*, 1971, *13*, 363-368.

Wheatley, D., & Unwin, A. *The algorithm writer's guide.* London, England: Longmans, 1972.

Wright, P., & Reid, F. Some alternatives to prose for expressing the outcomes of complex contingencies. *Journal of Applied Psychology*, 1973, *57*, 160-166.

## TYPOGRAPHY/GRAPHICS

Baker, S. *Visual persuasion: The effect of pictures on the subconscious.* New York: McGraw-Hill Co., Inc., 1961.

Barry, J. When posters get political. *Advertising and Sales Promotion*, 1968, *16*, 24-27.

Biggs, J. *Basic typography.* New York: Watson-Guptill Publications, 1968.

Blaiwes, A. Formats for presenting procedural instructions. *Journal of Applied Psychology*, 1974, *59*, 683-697.

Borg, W. & Schuller, C. Detail and background in audiovisual lessons and their effect on learners. *Educational Communication and Technology*, 1979, *27*, 31-38.

Bowman, W. *Graphic communication.* New York: John Wiley & Sons, Inc., 1968.

Chermayeff, I. Attitudes in design. *Direct Communication*, 1967, *53*, 24-30.

Chizlett, C. An analytical approach to problems of graphic design. *Print in Britain*, 1967, *15*, 18-19.

Coles, P. & Foster, J. Typographic cueing as an aid to learning from typewritten text. *Programmed Learning and Educational Technology*, 1975, *12*, 102.

Croxton, F. & Stein, H. Graphic comparison by bars, squares, circles and cubes. *Journal of the American Statistical Association*, 1932, *27*, 54-60.

Culbertson, H. & Powers, R. A study of graph comprehension difficulties. *Audio Visual Communication Review*, 1959, *7*, 97-100.

Dair, C. *Design with Type.* Toronto: University of Toronto Press, 1974.

Dwyer, F. Color as an instructional variable. *Audio Visual Communication Review*, 1971, *19*, 399-416.

Dwyer, F. *A guide for improving visualized instruction.* State College, Pa.: State College Learning Services, 1972.

Feliciano, G., Powers, R. & Kearl, B. The presentation of statistical information. *Audio Visual Communication Review*, 1963, *11*, 32-39.

Gregory, M. & Poulton, E. Even versus uneven right-hand margins and rate of comprehension in reading. *Ergonomics*, 1970, *13*, 427.

Guidry, N. & Frye, K. *Graphic communication in science.* National Science Teachers Association, 1968.

Hartley, J. & Burnhill, P. Contemporary issues in typographic research. *Programmed Learning and Educational Technology*, 1975, *12*, 73.

Hartley, J. & Burnhill, P. Understanding instructional text: Typography, layout, and design. In M. Howe (Ed.), *Adult learning: Psychological research and applications.* New York: John Wiley & Sons, 1977.

Hartley, J., Fraser, S. & Burnhill, P. A selected bibliography of typographical research relevant to the production of instructional materials. *Audio Visual Communication Review*, 1974, *22*, 181-190.

Hofmann, A. *Graphic design manual.* Reinhold Publishing Corporation, 1965.

Hymes, D. *Production in advertising and the graphic arts.* New York: Henry Holt and Company, 1958.

Kamman, R. The comprehensibility of printed instructions and the flowchart alternative. *Human Factors*, 1975, *17*, 90-113.

Landa, V. *Algorithmization in learning and instruction.* Englewood Cliffs, N.J.: Prentice Hall, Inc., 1974.

Levin, J., & Lesgold, A. On pictures in prose. *Educational Communication and Technology*, 1978, *26*, 233-243.

Lewis, B., Horabin, I., & Gane, C. *Case studies in the use of algorithms.* New York: Pergamon Press, 1973.

Macdonald-Ross, M. Graphics in text. In L. Shulman (Ed.), *Review of research in education* (Vol. 5). Itasca, Ill.: F.E. Peacock Publishers, Inc., 1978.

Macdonald-Ross, M. & Smith, E. *Bibliography of textual communication.* (Monograph No. 3). Milton Keynes, U.K.: Institute of Educational Technology, 1973.

Macdonald-Ross, M. & Smith, E. *Graphics in text: A bibliography.* (Monograph No. 6). Milton Keynes, U.K.: Institute of Educational Technology, 1977.

Nelson, R. *Publication design.* Dubuque, Iowa: Wm. C. Brown Co. Publishers, 1972.

Paterson, D. & Tinker, M. *How to make type readable.* New York: Harpers, 1940.

Poulton, E. Rate of comprehension of an existing teleprinter output and of possible alternatives. *Journal of Applied Psychology*, 1968, *52*, 16-21.

Rickards, J. & August, G. Generative underlining strategies in prose recall. *Journal of Educational Psychology*, 1975, *67*, 860.

Schmid, C. *Handbook of graphic communication.* New York: The Ronald Press Company, 1954.

Scott, R. *Design fundamentals.* New York: McGraw-Hill Co., Inc., 1951.

Sherbow, B. *Making type work.* New York: The Century Company, 1916.

Simon, O. *Introduction to typography.* London: Faber and Faber, 1963.

Spencer, H. *The visible word.* London: Lund Humphries, 1969.

Spencer, H., Reynolds, L. & Coe, B. *The effects of image degradation and background noise on the legibility of text and numerals in four different typefaces.* London: Royal College of Art, 1975.

Tinker, M. *Legibility of print.* Ames, Iowa: Iowa State University, 1963.

Tinker, M. *Bases for effective reading.* Minneapolis: University of Minnesota Press, 1965.

Tinker, M. Experimental studies on the legibility of print--An annotated bibliography. *Reading Research Quarterly*, 1966, *1*, 68-118.

Tinker, M. & Paterson, D. Influence of type form on speed of reading. *Journal of Applied Psychology*, 1928, *12*, 359-368.

Turnbull, A. & Baird, R. *The graphics of communication* (3rd Edition). New York: Holt, Rinehart and Winston, 1975.

Wason, P. The drafting of rules. *The New Law Journal*, 1968, *118*, 548-549.

Wheatley, D. & Unwin, A. *The algorithm writer's guide.* London: Longmans, 1972.

Wright, P. Using tabulated information. *Ergonomics*, 1968, *11*, 331-343.

Wright, P. Explicit and implicit tabulation formats. *Ergonomics*, 1972, *15*, 175-187.

Wright, P. Forms of complaint. *New Behaviour*, 1975, *1*, 206-209.

Wright, P. Presenting technical information: A survey of research findings. *Instructional Science*, 1977, *6*, 93-134.

Wright, P. *Is legal jargon a restrictive practice?* Paper presented at a meeting of the Social Science Research Council Law and Psychology Seminar Group, Oxford, England, 1978.

Wright, P. & Fox, K. Presenting information in tables. *Applied Ergonomics*, 1970, *1*, 234-242.

Wright, P. & Reid, F. Some alternatives to prose for expressing the outcomes of complex contingencies. *Journal of Applied Psychology*, 1973, *57*, 160-166.

Young, F. *Advertising layout.* New York: Covici, Frisde, Publishers, 1928.

Anderson, J.R. *Language, memory and thought.* Hillsdale, N.J.: Erlbaum, 1976.

Anderson, J. & Reder, L. Negative judgments in and about semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 664-681.

Bever, T.G. The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognition and the development of language.* New York: Wiley, 1970.

Bourne, L.E. Learning and utilization of conceptual ideas. In B. Kleinmuntz (Ed.), *Concepts and the structure of memory.* New York: Wiley, 1967.

Braine, M.D.S. Children's first word combinations. *Monographs of the Society for Research in Child Development*, 1976, *41* (1, Serial No. 164).

Braine, M.D.S. On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 1978, *85*, 1-21.

Bruner, J.S., Goodnow, J.J.,& Austin, G.A. *A study of thinking.* New York: Wiley, 1956.

Carpenter, P. & Just, M.A. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 1975, *82*, 45-73.

Charrow, R. & Charrow, V. Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review*, 1979, *79*, 1306-1374.

Charrow, V. & Charrow, R. *The comprehension of standard jury instructions: A psycholinguistic approach.* Arlington, Va: Center for Applied Linguistics, 1978.

Chase, W.G. & Clark, H.H. Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley, 1972.

Clark, H.H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, *76*, 387-404.

Clark, H.H. Difficulties people have in answering the question "Where is it?" *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 265-277.

Clark H.H. Semantics and comprehension. In T.A. Sebeck (Ed.), *Current trends in linguistics* (Vol. 1). The Hague: Mouton, 1973.

Clark, H.H. & Card, S.K. The role of semantics in remembering comparative sentences. *Journal of Experimental Psychology*, 1969, *82*, 545-553.

Clark, H.H. & Chase, W.G. On the process of comparing sentences against pictures. *Cognitive Psychology*, 1972, *3*, 472-517.

Collins, A.M. & Quillian, M.R. Does category size affect categorization time? *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 432-438.

Fillmore, C. Case for case. In E. Bach & R. Harms, (Eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart, & Winston, 1968.

Fodor, J.A., Bever, T.G., & Garrett, M.F. *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill, 1974.

Gough, P.B. Grammatical transformations and speed of understanding. *Journal of Verbal Learning & Verbal Behavior*, 1965, *5*, 107-111.

Gough, P.B. The verification of sentences: The effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 492-496.

Greene, J.M. The semantic function of negatives and passives. *British Journal of Psychology*, 1970, *22*, 14-27.

Griggs, R.A. & Shea, S.L. Integrating verbal quantitative information in linear orderings. *Memory and Cognition*, 1977, *5*, 287-291.

Haygood, R.C. & Bourne, L.E. Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review*, 1965, *72*, 175-195.

Horabin, I. & Lewis, B. *Algorithms*. Charles Town, W. Va.: Ivan Horabin, 1974.

Hunt, E.B. & Hovland, C.I. Order of consideration of different types of concepts. *Journal of Experimental Psychology*, 1960, *59*, 220-225.

Johnson, N.F. The psychological reality of phrase structure rules. *Journal of Verbal Learning & Verbal Behavior*, 1965, *4*, 469-475.

Johnson, N.F. Sequential verbal behavior. In T.R. Dixon & D.L. Horton (Eds.), *Verbal Behavior and General Behavior Theory*. Englewood Cliffs: Prentice Hall, 1968.

Johnson-Laird, P.N. The interpretation of quantified sentences. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.), *Advances in Psycholinguistics*. Amsterdam: North-Holland, 1970.

Johnson-Laird, P.N. Linguistic complexity and insight into a deductive problem. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.), *Advances in Psycholinguistics*. Amsterdam: North-Holland, 1970.

Johnson-Laird, P.N. Models of deduction. In R. Falmagne (Ed.), *Reasoning: Representation and processes in children and adults*. Hillsdale, N.J.: Erlbaum, 1975.

Johnson-Laird, P.N. & Tagart, J. How implication is understood. *American Journal of Psychology*, 1969, *82*, 367-373.

Jones, S. Instructions, self-instructions and performances. *Quarterly Journal of Experimental Psychology*, 1968, *20*, 74-78.

Just, M.A. Comprehending quantified sentences: The relation between sentence-pictures and semantic memory verification. *Cognitive Psychology*, 1974, *6*, 216-236.

Just, M. & Carpenter, P. Comprehension of negatives with quantification. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 244-253.

Just, M. & Clark, H.H. Drawing inferences from the presuppositions and implications of affirmative and negative sentences. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 21-31.

Katz, J.J. *Semantic theory*. New York: Harper and Row, 1972.

Kintsch, W. *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

Lakoff, G. Linguistics and natural logic. *Synthese*, 1970, *22*, 151-271.

Legrenzi, P. Relations between language and reasoning about deductive rules. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam: North-Holland, 1970.

Loftus, E.G. Category dominance, instance dominance, and categorization time. *Journal of Experimental Psychology*, 1973, *97*, 70-74.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning & Verbal Behavior*, 1963, *2*, 346-351.

Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, *63*, 81-97.

Miller, G.A. Some psychological studies of grammar. *American Psychologist*, 1962, *17*, 748-762.

Miller, G.A. & McKean, K. A chronometric study of some relations between sentences. *Quarterly Journal of Experimental Psychology*, 1964, *16*, 297-308.

Morris, J.O. *Make yourself clear!* New York: McGraw-Hill, 1972.

Neisser, U. & Weene, P. Hierarchies in concept attainment. *Journal of Experimental Psychology*, 1962, *64*, 640-645.

Perfetti, C.A. Lexical density and phrase structure depth as variables in sentence retention. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 719-724.

Quine, W.V.O. *Elementary logic.* New York: Harper & Row, 1965.

Rips, L.J., Smith, E.E., & Shoben, E.J. Semantic composition in sentence verification. *Journal of Verbal Learning & Verbal Behavior*, 1978, *17*, 375-402.

Rohrman, N.L. The role of syntactic structure in the recall of English nominalizations. *Journal of Verbal Learning & Verbal Behavior*, 1968, *7*, 904-912.

Rose, A.M. *An information-processing approach to performance assessment.* (Final report). Washington, D.C.: American Institutes for Research, November 1978.

Rose, A.M. & Cox, A. *Following instructions.* (Final report). Washington, D.C.: American Institutes for Research, 1980 (in preparation).

Sherman, M. Adjectival negation and the comprehension of multiply-negated sentences. *Journal of Verbal Learning & Verbal Behavior*, 1976, *15*, 143-157.

Sherman, M.A. Bound to be easier? The negative prefix and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 76-84.

Slobin, D.I. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 219-227.

Smith, E.E., Shoben, E.J., & Rips, L.J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychology Review*, 1974, *81*, 214-241.

Trabasso, T. Mental operations in language comprehension. In J.B. Carroll & R.D. Freedle (Eds.), *Language comprehension and the acquisition of knowledge.* Washington, D.C.: V.H. Winston, 1972.

Trabasso, T., Rollins, H., & Shaughnessey, E. Storage and verification shapes in processing concepts. *Cognitive Psychology*, 1971, *2*, 231-289.

Wales, R.J. & Grieve, R. What is so difficult about negation? *Perception and Psychophysics*, 1969, *6*, 327-332.

Wason, P.C. *Psychological aspects of negation.* London: Communications Research Centre, University College, 1962.

Wason, P.C. The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 1965, *4*, 7-11.

Wason, P.C. The processing of positive and negative information. *Quantitative Journal of Experimental Psychology*, 1959, *11*, 92-107.

Wason, P.C. Psychological aspects of inference. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.), *Advances in psycholinguistics.* Amsterdam: North-Holland, 1970.

Wason, P.C. Response to affirmative and negative binary statements. *British Journal of Psychology*, 1961, *52*, 133-142.

Wason, P.C. & Johnson-Laird, P.N. *Psychology of reasoning: Structure and content.* Cambridge, Mass.: Harvard University Press, 1972.

Wason, P.C. & Jones, S. Negatives: Denotation and connotation. *British Journal of Psychology*, 1963, *54*, 299-307.

Yngve, V.H. A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 1960, *104*, 444-466.